DOCUMENT RESUME

ED 146 204	95	TM 006 584

AUTHOR	Lunneborg, Clifford E.
TITLE	Interactive Assessment of Differences in Cognition. Final Report.
INSTITUTION	Washington Univ., Seattle. Dept. of Psychology.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Apr 77
GRANT	NIE-G-74-0104
NOTE	163p.; Appendixes C and D (Behavior Research Methods and Implementation) of the original document are copyrighted and therefore not included; not available in hard copy due to marginal legibility of original document; Appendix I is also available as ED 138 618
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Cognitive Ability; *Cognitive Processes; Computer Oriented Programs; *Computer Programs; *Individual Differences; Information Processing; Input Output; *Measurement Techniques; *On Line Systems; Phonotape Recordings; Programing Languages; *Psychological Testing; Research Design; Systems Development; Tape Recordings

ABSTRACT

An interactive assessment system which was built around an existing computer controlled testing facility in the Department of Psychology at the University of Washington is described. The Extensible Multiprograming System for experimental psychology (EMPP) is problem oriented software developed to simplify the programing of psychological experiments. The system is capable of controlling several experimental stations where the same experiment can be performed synchronously or asynchronously at each of the stations, or where stations may be totally independent. This document also includes progress reports on the development and operation of the system, conference papers outlining the system and its language (PL/E), a discussion of the problems involved in designing an on line speech input terminal for the system, and a technical report on the study of individual differences in cognition. In addition, there are reports on choice reaction time and its role in ability measurement and on the information processing correlates of the Raven Progressive Matrices Test. (Author/MV)

****
*   Documents acquired by ERIC include many informal unpublished   *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available   *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not   *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original.   *
****

ED146204

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

FR

Final Report

Grant NIE-G-74-0104: Interactive Assessment of Differences in Cognition

Clifford E. Lunneborg
Principal Investigator

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

TM006 584

Department of Psychology

University of Washington

April 1977

BEST COPY AVAILABLE

Final Report

Grant NIE-G-74-0104: Interactive Assessment of Differences in Cognition

Clifford E. Lunneborg, Principal Investigator

The grant project period ended 30 June 1976 and the following report with its appendices summarizes our accomplishments over the term of the grant.

Progress reports covering the initial (1 September 1974 to 28 February 1975) and middle (1 March 1975 to 31 October 1975) portions of the grant period have been submitted and are incorporated here as Appendices A and B respectively. In keeping with the major thrust of the grant, the final portion of the project period focussed on the refinement of EMPP, the computer controlled system for interactive assessment of cognitive performance.

The EMPP hardware configuration was not changed in the final period but the software, particularly the higher level language used for the description by researchers of experimental tasks, was considerably advanced.

While no new data gathering efforts were mounted a number of analyses of earlier studies were undertaken and several reports completed. These will be referenced below.

System Development

Under this grant federal funds were sought and expended for the purpose of developing and testing an interactively capable integration of hardware and software permitting the computer control of cognitive studies. Because the system is intended for use in studies of individual differences in cognition it was important that it have very fast response times while tracking individual subjects and tailoring stimulus sequences to the individual's responses.

Although there were a number of computer controlled systems already in use in laboratories of cognitive psychology these were oriented either towards a single subject or towards yoking a set of subjects to a common experimental paradigm. Efficient prosecution of individual difference studies demanded that we be able to assess several subjects simultaneously but had to permit the assessment of any individual as a function of that individual's particular cognitive abilities.

To achieve the desired system required that we simultaneously work on the improvement of an initial computer-controlled laboratory and on the development of interactive software. An important side-condition on this software development was that the product had to be usable by psychologists with only minimal experience in computer programming.

We believe we have met these goals rather effectively. Development of the system is nicely summarized in a series of papers prepared principally by Kenneth J. Burkhardt, Jr., now Assistant Professor of Electrical Engineering, Rutgers University. Dr. Burkhardt was responsible for both hardware and software system design. These papers make up Appendices C, D, and E to this report.

Appendix C, published in Behavior Research Methods and Instrumentation, develops fully the goals of an asynchronous, multi-paradigm system and outlines the software structure which was first implemented. EMPP, as deployed for use in 1975 provided for the asynchronous operation of our eight subject stations, permitting multiple experiments to be underway simultaneously. The EMPP software was written to be extensible and was placed in operation to meet our data collection needs effectively although refinements in ease of hardware control were known to be desirable and planned. As the later System Operation section in this report will bear testimony to, this first phase system permitted our laboratory to accomplish a considerable amount of human subject evaluation.

A second system development paper, Appendix D which is also a Behavior Research Methods and Instrumentation publication, describes an extremely important hardware development, one which was essential if EMPP were to provide the measurement resolution necessary to study differences in human response to very short duration stimuli as indicators of information processing efficiency. The "intelligent station controller", designed by Burkhardt and realized for us by IKON Corporation, permits the very accurate timing of the onset and offset to our subject stations of a wide range of visual displays. We can now control onset times to within 1 msec., an achievement the more remarkable when it is noted that this holds independently for the eight displays controlled.

The performance of Burkhardt's station controller is a function of the evolution of EMPP software as well as of the controller architecture itself. In this refinement, a higher level language, PL/E, was provided for EMPP. PL/E was designed, specifically, to be very easily employed in experiment design by psychologists untrained in computer programming and relatively unsophisticated in computer-controlled experimentation. Appendix D describes the CRT control commands presently implemented in PL/E as well as the controller configuration.

Appendix E, The evolution of a higher level language for an on-line minicomputer system, addresses the development of PL/E more broadly. The primitive compiler capability of the earlier EMPP system has been markedly improved. In addition to making minor changes to improve program readability and compile and run time diagnostics; providing more useful support for algebraic and conditional statements; and adding graphics capabilities and more sophisticated loop control, the second stage of EMPP refinement defined the syntax and semantics of our "experiment-describing language", PL/E. Two major goals were to guide development: (1) PL/E had to be a significantly easier language in which to program and debug cognitive studies than our earlier version. (2) PL/E had to have a structure which would accomodate extensibility; neither the hardware configuration of our individual differences laboratory nor the range of experimental paradigms to be investigated can be considered fixed or completely foreseeable. The first goal has been met, the second can only be evaluated over time.

One of the extensions considered important is to make the EMPP system capable of recognizing and responding to a limited spoken English vocabulary. Characteristically humans signal that they have completed information processing by vocalizing a response. The key-press format to responding imposes an additional, somewhat alien aspect to the task. While obtaining response times to verbal responses is certainly now possible, the fully interactive assessment system needs to be able to branch on the content of the verbal response as well as on its latency. Initial developments of this kind will naturally embrace the computer-recognition of a very small vocabulary: "Yes", "No", "Left", "Right", "First", "Second", "Larger", "Smaller", "Same", "Different", for example, constitutes a set of verbal responses which would permit the natural exploration of a wide range of information processing tasks. Such a capability does not appear to be presently economically attainable. In Appendix F, Dr. Burkhardt reviews the state of the art in computer-controlled speech recognition and describes the potential organization of such processing equipment in an individual differences oriented laboratory of cognitive psychology.

In summary, the EMPP system as presently configured and supported by the PL/E language provides highly effective support for the continued study of cognitive processes. The system, indeed, is presently highly saturated with research studies. Orientation of these continuing studies will be summarized below.

System Operation

Under this grant and under a companion grant to the Department of Psychology from the National Institute of Mental Health ("Individual Differences in Cognition", Earl B. Hunt, principal investigator) we have accomplished considerable by way of exploring cognitive performance from the point of view of modern information processing theories. Characteristically, we have undertaken what we term "Marathon" studies; studies in which a panel of 50-75 subjects participate in a variety of selected psychometric and laboratory assessments. The need to obtain a multiplicity of measures for groups of this size in the relatively short time period that such groups may be held together for research work highlights the justification for the EMPP system with its eight independently controllable subject stations.

Professor Hunt, co-principal investigator, reviews in the accompanying Appendix G the substantive studies completed under this grant and under the NIMH grant in its first four years, paying particular attention to the series of studies which relate information processing efficiencies to verbal intelligence. This review admirably summarizes our work.

Two, more limited studies by the principal investigator make up Appendices H and I. Choice reaction time: what role in ability measurement? will appear in Applied Psychological Measurement and summarizes a series of studies in which we related psychometric abilities to response latencies in simple choice studies. The problem pointed out in the conclusion to that paper, the need to be able to determine the reliability of laboratory assessments of information processing parameters, has led to one of the most recent extensions of EMPP. John Palmer of our laboratory has implemented an option which permits monitoring by the controlling experimental program of a moving average and moving index of variability for, as an example, choice reaction time. The goal, of course, is to obtain "locally reliable" assessments, ones that are maximally stable for the individuals being assessed.

In Information-processing correlates of the Raven progressive matrices test, which was presented as part of symposia on individual differences in cognition at the annual meetings of the American Educational Research Association and, in a slightly different form, of the Western Psychological Association, we begin to address the question of information processing and non-verbal intelligence. The apparent strong differentiation between the information processing correlates of psychometric measures of speeded versus power aspects of intellectual performance (as well as the verbal/non-verbal distinction) is one which can only be fully explored with a system like EMPP. Both accurate timing of stimulus and response events and the selection of problem difficulty (power) to match the individual subject's characteristics are essential to further work.

Continuation of Effort

The research program in the assessment of individual differences in cognition is continuing as a fully active one in the Washington laboratory. The EMPP system, developed under this grant, is expected to be fully committed to additional federally supported work over the next several years. At present the laboratory is conducting a series of studies under two grants. A continuation of the NIMH grant for four years will permit us to explore further individual differences in memory access characteristics and in the interactions between process and structure in problem solving. A two year grant from the National Institute of Education will carry our information processing approach to the study of the development of different degrees of reading competence. Professor Hunt is principal investigator for both of these grants.

Additional support is pending and we feel it is clear that NIE development of EMPP has provided our laboratory with a unique capability for furthering our understanding of man's intellectual abilities.

Department of Psychology
University of Washington
Seattle, Washington 98195

April 1975

Progress Report: 1 September 1974 - 28 February 1975

Grant NIE-G-74-0104: Interactive Assessment of Differences in Cognition

Principle Investigator: Clifford E. Lunneborg

Progress over the first six months of the grant period is summarized below. Although treated topically, work under the several headings has been going on in parallel.

1. System Development (Equipment Design and Acquisition). A major goal of the funded project is to build an interactive assessment system around an existing computer controlled testing facility in the Department of Psychology. This was to have two aspects; upgrading the equipment to handle, interactively, eight test stations and developing the software to handle the required interactive stimulus presentation and response recording. Progress on the equipment development is as follows:

a. The dual disc storage system and controller was acquired under bid from Ball Computer Products and installed late in February. Availability of this storage feature will permit implementation of the desired adaptive testing capability of administering stimuli from different experiments ("tests") simultaneously.

b. The four additional Tektronix CRT displays have been ordered for Spring delivery. Because of our requirement for very fast decay phosphors, a relatively long lead time is required for delivery. The scopes will be on hand, however, for interfacing with the new controller (c. below) and for the Summer testing programs. To reduce the likelihood of losing testing capability through scope failure a standby fifth scope has also been ordered.

c. A new station controller has been designed and a call for bids issued. Bids will be received mid-April and the controller (and keyboard stations) installed for the Summer. To refresh the eight scopes, each with a potentially different stimulus configuration under adaptive testing trials, cannot be accomplished by simple extension of the present four station controller. The new controller will permit a form of "local display generation" as a processor with limited storage, separate from the experiment driving Nova, will be dedicated to stimulus development and refreshment. This new design will be fully capable of expansion to 16 or more stations.

d. Microphones have been requested and will be installed in the existing four stations for Spring testing and included in the development of the four additional stations. Through the early Summer these will function solely as voice keys, operating through the present keyboard response circuitry.

e. Specification of characteristics for the analog to digital (A/D) signal convertors necessary to permit limited voice recognition has been undertaken. Each of the eight stations will be equipped with a separate A/D channel permitting essentially real-time processing of voice responses. Specifications will be issued to potential bidders in May and software for limited use of the convertors developed for August testing. This initial use will be with "Yes-No" vocal responses.

f. Audio tape playback equipment, permitting stimuli to be positioned and cued by computer, will also be acquired for August testing. (Although not funded by this grant, audio tape research in our laboratory this winter has centered on developing a capability for the computer generation of two channel "speech" tapes. This has been successfully accomplished and tapes produced in this way will be used in our testing programs.)

2. System Development (Programming). Although we had to program specific experiments as described in section 3 below, our software efforts have focused on developing a programming system which we refer to as "EMPP".

EMPP, an Extensible MultiProgramming system for experimental Psychology, is a problem oriented language (POL) developed to simplify programming of psychological experiments. The system is capable of controlling a number of experimental stations (eight, presently) where the same experiment can be performed synchronously or asynchronously at each of the stations or where stations may be totally independent in the sense that up to eight unrelated experiments can be performed simultaneously. The EMPP language is not directly related to conventional (data processing) programming languages and was developed primarily for the class of applications encountered in our laboratory.

Although there have been several attempts to develop specific POL's for use in computer controlled psychological laboratories, no one language has gained universal acceptance. This lack of a standard language is a function of economics, differences in applications and differences in physical hardware between laboratories. Rather than attempt to modify one of the existing languages to operate in our own environment, a decision was made to develop a language that would make use of our specific hardware capabilities and our need for "adaptive" shifts in presentations to simplify programming of experiments. The language is general enough, however, to be implemented on other similar systems with a minimum of difficulty.

The primary design goal was to produce a system with the following characteristics;

A. Simplicity - Psychologists with minimal programming experience should be capable of writing experimental programs in a short period of time.

B. Hardware Control Capability - It should be easy to display data on CRT's, record responses from keyboards and make use of other system hardware facilities.

C. Asynchronous Operations - It should be possible to operate experimental stations independently.

D. Multiplicity of Experiments - The system should be capable of simultaneously performing more than one experiment.

E. Extensibility - It should be possible to add new software and hardware facilities to an existing system.

Extensibility is perhaps the most important feature of this system. One of the major problems with existing systems is that situations often arise where a particular function cannot be performed by the existing language, (e.g., permute a list). The experiment programmer is then forced to manipulate the language in an attempt to implement the desired function. In some situations this is time consuming, in others the function simply can't be implemented.

It is easiest to describe a POL such as EMPP by describing the operating system under which the language operates. The EMPP operating system consists of four parts;

A. Subject Program - A series of commands written in the EMPP language that control the execution of individual experiments. An individual subject program may control one or more stations.

B. The Interpretor - The master system scheduling and control program. Schedules the execution of subject commands and passes control to the experiment macros.

C. Experiment Macros - A series of Macro subroutines (approximately 25) written in NOVA machine code that perform operations requested by the subject programs.

D. System Subroutines - A series of 25 general purpose subroutines used to simplify the coding of experiment macros. System subroutines typically perform operations required by several experiment macros (e.g., argument address calculations).

In a sense, the EMPP system involves the creation of a pseudo computer where the instruction set of the pseudo machine is a higher level language defined by the experiment macros. Each experiment macro represents an operation or series of operations to be performed for a particular subject. For example, the following would be required to display a number and record a keyboard response:

```
DISPLAY I
RECORD RESPONSE
  PRESP (console keys to be activated)
  RESP  (actual response)
  RT    (reaction time)
```

EMPP contains experiment macros for performing external control functions (delay, activate switches in consoles, etc.), internal control functions (branch, "DO" loop control, etc.), arithmetic computation (multiply, add, random number generation, etc.) and simple input output (type, display, plot, etc.).

The EMPP system has been coded and thoroughly debugged. The system supports all hardware presently connected to the system. Several experiments have been written and one was run as a part of the Winter testing.

We have, for the most part, been successfull in meeting our design goals. Experiment writing, however, is still fairly complex for non-programmers. To alleviate this problem we are in the process of writing a Compiler for the EMPP language which will greatly simplify the programming of experiments.

3. Data collection. Under the joint sponsorship of this grant and an NIMH funded project (Individual Differences in Cognition) both group psychometric data and cognitive laboratory performance data are being collected on a new panel of 60 university freshmen. Picked to be representative of the freshman population (on high school based measures of verbal and quantitative aptitudes) this panel will be extensively studied through the entire 1974-75 academic year.

During the period covered by this report, psychometric and laboratory testing covered the following areas.

Group Psychometric Data: Tests administered included:

a. Readministration of most of the WPC battery administered in high school (includes tests of verbal, quantitative, and spatial ability);

b. Such advanced-level cognitive ability tests as the Terman Concept Mastery, Watson-Gleser Critical Thinking, (pseudo) Miller Analogies, and the Advanced Set of Raven Progressive Matrices tests;

c. A series of short primary mental abilities tests -- verbal and symbolic reasoning, two and three dimensional spatial visualization, numerical computation and reasoning, perceptual speed and accuracy, maze tracing, code learning and code recall;

d. A set of "cognitive styles" tests -- hidden figures (field dependence/independence), smallest and largest instances (category width setting), role repertory and similarities and dissimilarities (category flexibility);

e. An information processing bettery (mark making, letter checking, and grammatical reasoning) adapted from a University of Michigan human performance project; and

f. A set of interest measures (Allport-Vernon, Vocational Preference, and Vocational Interest).

Laboratory Testing, Laboratory set-ups, for the most part computer controlled, included:

a. Binary Choice Reaction Time -- reaction times for visual cue triggered left and right key responses with data obtained for both one hand and two hand responding;

b. Posner Visual Field Task -- the Posner name and physical identity task with stimuli selectively presented to left or right visual fields;

c. Dichotic Listening Accuracy -- both speech and non-speech stimuli presented dichotically with varying within trial inter-stimulus intervals;

d. Memory Span Study -- recognition accuracy for long term and short term memory sets intermixed with short term set size systematically varied;

e. Temporal Order Judgement -- accuracy of judgement of order of stimuli presented to left, right, or both visual fields with varying ISI;

f. Raven Matrix, Set A -- response times for individual item solutions obtained,

g. Problem Solving Task -- true/false questions of systematically varied complexity based on simple scenario committed to memory a week earlier; and

h. Delayed Auditory Feedback -- Ss reading feedback with delay errors and reading speed to be analyzed.

Analyses of these data have only begun (see Reports section below) and further testing of this panel will involve individual clinical assessment of intelligence and problem solving strategies. Based on the analyses of this study two new subject panels will be recruited for Spring and Summer study -- a group of older (40-45 years) returning students and a wide ability range group of high school juniors.

4. Reports. The accompanying technical report, "Individual differences in memory and information processing", was presented to a symposium on individual differences in cognitive style held during the American Psychological Association (APA) meetings in New Orleans. It summarizes the work conducted in our laboratories immediately preceding that outlined above.

Analyses of results from the present panel studies will be presented at the APA Chicago meetings this Summer. Enclosed is an advance abstract, "Choice reaction time and psychometric performance revisited", of the accepted contribution. Associate Investigator Earl B. Hunt will also be reporting on work in our laboratory to a Division 3 meeting of APA.

5. Project Staff. During the six months covered by this report, staff paid by the grant have included:

a. Pre-doctoral research associate, 100% time, Kenneth Burkhardt -- a computer scientist, Burkhardt is responsible for equipment acquisition and systems programming and will be employed as a post-doctoral research associate part-time when he finishes his doctorate;

b. Graduate research assistant, 50% time, Edward Cimler -- a graduate student in cognitive psychology, Cimler is responsible for preparing and analyzing laboratory testing studies;

c. Secretary, one-third to 50% time, Maureen Schoener -- shared with the NIMH grant and Cognitive Psychology, Schoener has been responsible for subject recruitment and management as well as project secretarial and clerical details; and

d. Programmers (hourly) 40-80% time, Barbara Lewis, Bruce Geilhood, John Palmer, and Brenda Zabinsky -- under Burkhardt's direction these hourly programmers have been responsible for programming individual computer controlled experiments and for contributing to the elaboration of the systems programming.

With the exception of the change in status for Burkhardt, the project staff is expected to remain at the same level.

Department of Psychology
University of Washington
NI-25
Seattle, Washington 98195

8 December 1975

Progress Report: 1 March 1975 - 31 October 1975

Grant: NIE-G-74-0104 "Interactive Assessment of Differences in Cognition"

Principal Investigator: Clifford E. Lunneborg

Progress over this second part of the Grant period is summarized, topically, below.

1. System Development (Equipment Design and Acquisition) A major goal of the project is to build an interactive assessment system around an existing computer controlled laboratory facility in the Department of Psychology. This has two aspects; upgrading equipment to sample, interactively, eight test systems and developing the softwares to sample the required stimulus presentation and response recording. Progress on equipment has been as follows:

a. A new station controller has been designed, fabricated, and installed by Ikon. This controller includes memory or storage for each of the eight stations external to the memory associated with the controlling Nova. This permits CRT displays at each station to be independently Onset, Offset, Refreshed, Modified, or Replaced with only minimal demands on the experiment controlling Nova 830.

b. To insure compatibility and overcome earlier experienced mechanical and electronic failures the four pre-existing keyboards (response stations) have been replaced with units identical to the four new ones. All were built by Ikon and installed together with the station controller.

c. The additional fast Phospher Textromx CRT monitors have also been installed. The eight stations contemplated have now all been established with display scopes and response keyboards. The stations have all been operated simultaneously and independently in conjunction with the new controller and performed well. Although not put to operational use yet, each of the testing stations has now been equipped with re-designed and newly fabricated switching gear to permit computer control of "peripheral" input/output devices at each station: Voice-keys, slide projectors, audio playback and recording.

d. Development of our own specifications for the acquisition or fabrication of limited vocabulary voice recognition equipment has been delayed with the announcement of the commercial availability of relayed systems (e.g., for the control of machine tools.) Although presently available systems are expected to be too expensive to acquire, their promotion is evidence that there is now a commercially feasible technology. Dr. Burkhardt will explore this over the winter.

e. Deployment of several of our new testing paradigms on the system, simultaneously, would apparently be facilitated by expanding core memory for the Nova from 16 to 24 words. A decision on funding this expansion from this or the related NIMH grant under the direction of Professor E. B. Hunt will be made shortly.

f. Beyond items discussed here, no further equipment acquisition is planned nor required to meet our anticipated testing interests.

2. System Development (programming). The Extensible Multi Programming system for experimental Psychology (EMPP) outlined in the first progress report has now been almost fully developed. The operating system portion, EOS, is fully implemented and is functioning to control the eight testing stations in both the synchronous (all stations yoked) and asynchronous modes. In the asynchronous mode not only may the eight stations run through the same experiment at different rates but each of the eight stations may indeed be participating in a separate experiment, all at the same time.

Availability of the dual disc storage and related disc operating system acquired earlier has facilitated the development of EMPP. In particular, a compiler for the programming language has been developed and is now in use. Similarly, a very comprehensive set of input/output commands for use with the subject testing stations have been made available. Both improvements were designed with the psychologist with minimal programming experience in mind. The facility with which project staff have been able to write experimental control programs this autumn suggests these efforts have been largely successful.

The latest EMPP extension has been the development of a graphic display capability for the CRT stations. The control language promises to be fairly easy to use and this capability will be a valuable augmentation of our alphanumeric display generation system. EMPP documentation, necessary for full utilization of the system, is now available in the form of a Users Manual and a Systems Programmer Guide. The modular, extensible nature of EMPP permits programming attention to now be staffed to the development of a central structure for peripheral devices other than the CRTs and response keyboards.

3. Data collection. Testing of University freshmen undertaken during the first period of this project was continued through the spring term. The panel of 64 students were extensively tested in the laboratory and in group paper-and-pencil psychometric sessions. Some initial analyses of these data have now been communicated (see Reports Section below.)

Because of the need to obtain more widely representative cognitive performance data a group of 64 high school juniors were brought to the testing stations to complete an abbreviated (2-day) battery of laboratory and psychometric measures during the summer. These testing sessions, as with the earlier freshman ones, were jointly supported by this and an NIMH grant directed by Professor Hunt.

Pilot studies of older students, 45-55 years of age, were undertaken in the late spring. These studies have concentrated on determining the requirements for obtaining stable estimates of speed of processing parameters in a Posner-type task (detecting "identical" visual displays) and relating these to selected standard paper-and-pencil measures. Experience with these Ss has been particularly valuable in establishing more appropriate sizes for displays, lengths of interstimulus intervals, and subject orientation.

An outgrowth of this work with older Ss is expected to be the development of efficient paper-and-pencil forms of three of the more popular information processing tasks--Posner identity detection, Baddaley grammatical reasoning, and Shepherd-Metzler speed of rotation. These will be validated against laboratory performance in high school, University, and older Ss.

The present emphasis on completing the testing system together with the need to conduct further analyses of already collected data have resulted in scheduling no new testing studies this fall.

4. Reports. The principal investigator reported certain results from the university and high school samples mentioned above at the annual meeting of the American Psychological Association in Chicago. The accompanying technical report, "Choice reaction time and psychometric performance revisited", is based upon that report. A modified version has been submitted for Journal publication.

Dr. Kenneth Burkhardt, Research Associate responsible for design and development of our experiment-controlling programming language reported on this work at the fall Boulder conference on Use of Computers in Psychology. A copy of that paper, "EMPP: an extensible multiprogramming system for experimental psychology", accompanies this report. Publication of this description of our system will also be sought.

The principal investigator also attended an invitational conference on Computerized Intellective Testing held in Washington in June and jointly sponsored by ONR and the U.S. Civil Service Commission. Although the emphasis was on tailoring multiple-choice tests to respondents there was an awakening interest in the possible role of laboratory-like measures of informative processing speed in individual assessment.

5. Project staff. Project staff has remained essentially unchanged. Dr. Burkhardt has coordinated the work of grad hourly programmers; graduate research assistant Edward Cimler coordinated spring and summer group testing; and undergraduate research assistant Jean Bekey had responsibility for pilot work with older subjects.

Appendix C has been removed due to copyright restrictions.

The article, "EMPP: An Extensible Multiprogramming System for Experimental Psychology," is published in: Behavior Research Methods and Instrumentation; v8 n2 p239-44 1976.

Appendix D has been removed due to copyright restrictions.

The article, "An Intelligent Experimental Station Controller for Human Experimental Psychology," is published in: Behavior Research Methods and Instrumentation; v8 n4 p369-74 1976.

The Evolution of a Higher Level Language for an On-Line Minicomputer System

Kenneth J. Burkhardt, Jr.

Abstract

The purpose of this paper is to describe the use of PL/E, a loosely defined higher level language, in the implementation of an extensible multiprogramming system for experimental psychology (EMPP). Although this system was designed with a specific class of applications in mind (namely human experimental psychology), the resulting system and the methodology used in its development are applicable to a large number of real-time systems.

The paper consists of three sections: The first describes the environment in the experimental laboratory at the University of Washington and the nature of the problems that the EMPP system was designed to resolve. The second section describes the resulting system from the standpoint of the special purpose language and the structure of the operating system used in its implementation. The final section describes the methodology used in developing the EMPP system.

The Evolution of a Higher Level Language for an On-Line Minicomputer System

Kenneth J. Burkhardt, Jr.

One of the most common uses of minicomputers is in small to medium sized on-line, real-time systems applications. Minicomputers have been connected to chemical processes, nuclear accelerators, automatic x-ray machines and used in a variety of production and research environments. There are more than fifty minicomputers at the University of Washington alone connected to everything from a linear accelerator to monkeys in the primate laboratory.

In the early days of laboratory computer use there was limited concern for the cost and methodology of minicomputer programming. Typically a researcher would acquire a small computer and sit down and program it himself in a brute force fashion. Although the resulting system would often do the job, the major problem with brute force programming was that the researcher ended up spending more time worrying about the computer than performing his research. As the cost of acquiring small computers has decreased, the number of potential users has increased. The result has been that many potential users are totally unfamiliar with computer software and hardware concepts. Thus the need for developing more effective tools for software development has grown considerably.

The development of effective software for on-line computer systems can be greatly enhanced by the use of higher level languages. There have been many languages implemented on minicomputers. Most, however, are not well suited to a large number of real-time computer applications. The ISA has recommended a number of extensions to FORTRAN, however, the implementation of these extensions has not been uniform.[1] Even if it were, FORTRAN is certainly not the answer for applications that require character manipulation. ALGOL with its needs for recursive procedures and run time stacks is often too memory inefficient. Special purpose languages such as DEC's INDAC and IBM's PROSPERO have met with mixed success in the process control area, but are not applicable in many areas. Thus the majority of real-time minicomputer applications are still programmed in assembly language. This leads to an inordinately large effort being spent on both the development and the testing of minicomputer software.

During the past several years I have been interested in developing a technique to minimize the need for assembler programming of real-time minicomputer applications. PL/E is an attempt to specify a language and an operating system philosophy that can be utilized to simplify the implementation of a broad class of minicomputer applications. The PL/E language is not completely specified, but is being refined as experience with its use grows.

The purpose of this paper is to describe the use of a subset of PL/E in the implementation of an extensible multiprogramming system for experimental psychology (EMPP) at the University of Washington. Although this particular system was designed with a specific class of applications in mind (namely human experimental psychology), the resulting system and the methodology used in its development are applicable to a large number of real-time systems.

This paper consists of basically three sections: The first section discusses the environment in the experimental laboratory at the University of Washington and the nature of the problems that the EMPP system was designed to resolve. The second section describes the resulting system from the standpoint of the PL/E language and the structure of the operating system. The third section describes the methodology used in developing the EMPP system.

The Experimental Laboratory

Hardware Facilities

Figure 1 is a block diagram of the hardware facilities of the psychological testing laboratory at the University of Washington. The laboratory system is

Insert Figure 1 about here

controlled by a Data General NOVA 820 minicomputer equipped with 16K words of 16-bit, 800-nanosecond core memory. The NOVA has a primitive priority interrupt system, a programmed real-time clock, high and low speed direct memory access facilities and integer multiply and divide hardware. Programs and data are ~~stated~~ stored on either one of two 2.5 megabyte moving arm disks or on one of two LINC tape drives.

Communication with individual experimental subjects is through an intelligent experimental station controller. The station controller oversees the operation of eight independent experimental stations. Each station contains a CRT display and a console interface. The console interface allows the controller to recognize subject responses from special devices such as voice keys or telegraph keys or from one of eight keys located on the consoles. Eight bits of digital output are available for latching relays or solid state switches.

Programmer interface to the system is via an IMLAC graphics display, a teletype and a medium speed printer.

Experimental Environment

The use of the system facilities is a hands on operation in the literal sense of the term. There are no professional programmers, keypunchers or other support personnel. Experimental programming is performed either by the person who wishes to perform the experiment or one of his students.

A typical experiment involves presenting a stimulus at each of the experimental stations and then recording the subject's reaction time and response to the stimuli. For example, a simple motor skills test might involve presenting a figure in the subject's right or left visual field and then expecting either the rightmost or leftmost key on the console to be pressed. A more complex experiment might require some sort of adaptive or deductive testing. For example, a subject might be asked "Does 2 x 2 = 6?" If the subject answered the question correctly, the next question might be "Does 4 = Sqrt(16)?" If the subject answered incorrectly, the next question might be "Does 2 x 1 = 2?"

There is a wide range of complexity of experiments. Both of these sample tests are fairly simple in concept, however, adaptive or deductive tests can be quite complex.

The frequency of execution of an experimental program varies with the purpose of the experiment associated with it. "Standard" experiments may be given to a large number of subjects over an extended period of time. Other experiments function primarily as pilot studies in which case the experimental program may be run only a few times and then either modified or discarded. In terms of number of experimental programs developed in the laboratory the majority fall into the latter class. Thus there is a constant demand to be developing new experimental programs.

Software Tools

Prior to the development of the EMPP system it was not possible to rapidly develop new experimental programs. Experimental control programs were written strictly in assembly language with all the resulting inefficiencies. Several macro level I/O handlers were available, but they were quite primitive and fairly difficult to use. The major problem with assembler programming was the total lack of any unified operating system or any scheduling facilities. The only experiments that were developed were those that required all subjects to be performing the same operation at the same time. Under this mode of operation, if a particular subject took an inordinately long period of time to respond to a stimulus, all other subjects would experience corresponding delays in the presentation of new stimuli. It was not possible to program either deductive or adaptive experiments. Obviously the laboratory was not being used to its potential.

The EMPP System

The primary justification of the EMPP system was to simplify the development of experimental programs by providing a higher level language oriented towards human experimental psychology. Although there have been several attempts to develop specific problem oriented languages for computer controlled psychological testing laboratories (2, 3, 4), no one language has gained universal acceptance. From our standpoint the greatest drawback of existing psychological systems was their orientation towards a single subject. There were no facilities for parallelism control. A secondary drawback was the difficulty of learning many of the languages. SCAT (5), one of the more widely used psychological languages, uses a finite state notation technique to specify program control flow. Several of the other languages are merely macro extensions of assembly languages. Since the majority of programs in this system would not be written by programmers or engineers, the use of state notation or macro subroutines would be foreign to the average programmer. Several of the ALGOL-like languages appeared promising, but did not contain the necessary editing and character manipulation facilities. The "new language" would have to offer the mathematical power of FORTRAN or ALGOL together with a degree of readability that could be understood by nonprogrammers.

System Design Goals

The primary goal in designing the EMPP system was to produce a system with the following characteristics:

(1) Simplicity. Psychologists with minimal programming experience should be capable of writing experimental programs in a short period of time.

(2) Hardware Control Capability. It should be possible to display data on CRT's, record subject responses at keyboards and make use of other system hardware facilities.

(3) Asynchronous Operations. It should be possible to operate experimental stations independently.

(4) Extensibility. Since it is difficult (if not impossible) to predict all operations required by a general set of psychological experiments, it should be possible to add new software and hardware facilities to an existing system.

Structure of the EMPP System

The resulting system should be viewed as evolutionary. EMPP started with a basic set of primitive commands and has grown with the need for more complex operations. This ability to grow makes it difficult to rigidly define the EMPP system. Even though the system may change, however, the basic system concepts will remain the same.

The EMPP system consists of four program types. (1) Subject programs consist of a series of commands written in the PL/E language that control the execution of individual experiments. An individual subject program can control one or more experimental stations. (2) The master system scheduling and control program schedules the execution of subject program commands and passes control to the experiment macros. (3) The experiment macros are a series of macro subroutines written in NOVA machine language that perform operations requested by the subject programs. (4) The system subroutines are a series of general purpose subroutines used to simplify the coding of experiment macros. System subroutines typically perform operations required by several experiment macros (e.g. argument address calculations, list processing function, etc.).

In a sense, the EMPP software system can be viewed as a pseudo computer where the instruction set of the pseudo machine is the meta language defined by the experiment macros.

The relationship of subject programs to the other EMPP modules is depicted graphically in Figure 2. Each subject has a unique data storage area for recording experimental results. Data are recorded in these areas via subject program commands. The actual data transfers are physically performed by the experiment macros or the system subroutines. When a subject program issues a command, the scheduler causes control to be passed from the subject program to the appropriate experiment macro. The experiment macro either performs the requested operation or transfers control to a system subroutine. On completion of the requested operation control is returned to the subject program.

Insert Figure 2 about here

Although there are unique data areas for each subject, the system requires only a single copy of the subject program to be located in memory.

PL/E Language Description

A subject program consists of three sections:

(1) The Data Definition. Describes the type and scope of variables to be referenced by the experimental procedure.

(2) The Experimental Procedure. Specifies the operations to be performed during an experiment.

(3) The Format Description. Specifies data formats for all I/O operations.

Data Definition

PL/E variables can be described as global or local, integer or alphanumeric, and scalar or vector. Global variables are variables that can be accessed by all subjects in an experiment. Local variables are variables unique to each subject. For example, a global variable might refer to the number of trials in an experiment. A local variable might be the reaction time of a particular subject to a stimulus. Variables of any data type can be either scalar or vector (singly dimensioned array). The normal convention is to treat numeric variables as single word integers. Facilities exist, however, to process multiple precision variables. Alphanumeric variables are packed two characters per word and vary dynamically in length.

Writing Procedures

Procedures consist of three types of statements;

1) Command Statements
2) Assignment Statements
3) Conditional Statements

A command statement is a natural language-like directive such as

ADD C TO Y STORE IN Z

An assignment statement is an algebraic representation of one or more command statements. Assignment statements can use either symbolic abbreviations (e.g. +,-) or commands. For example,

Z = Y + C

is equivalent to

Z = ADD X to Y

Conditional statements are two tailed IF-THEN-ELSE expressions.

Command Types

There are five types of commands in PL/E: Arithmetic and string manipulation, program control, input/output control, real-time control and parallelism control.

Arithmetic and program control commands correspond to operations found in most algebraic languages. Commands exist for performing conventional arithmetic operations, logical operations and special arithmetic functions such as generation of random numbers and permutations of lists. String manipulation facilities include commands for performing string searches, string replacements, catenations, truncations, etc. for both fixed and variable length character strings. Program control commands include facilities for conditional and unconditional branches, subroutine branches and interactive loop control.

Input/Output Commands

Input/output commands allow I/O to any device to be rigidly defined or relatively free of fixed format constraints. The PL/E language accepts binary, octal or decimal numerics as well as fixed and variable length character strings. Special commands are available for positioning displays, loading and clearing CRT character memories and for starting and stopping display refresh operations.

Real-Time Control Commands

Real-time control commands allow a program to suspend a subject either as a function of time or as a function of an external event. For example, WAITRT causes the suspension of a subject until either a desired response is encountered at one of the consoles or a time limit expires.

Parallelism Control

PL/E programs are written as if only a single subject could participate. At run time the EMPP operating system determines which stations have subjects and causes a copy of the subject program to be executed at each of the appropriate stations. Through the use of FORK and JOIN instructions the programmer can specify whether a program segment will be executed once for all subjects or once by each subject. The "global" mode of execution is used primarily for initialization of the data base and for experimenter communications. The actual experiment generally takes place in the "local" mode.

A subject program is put in the local mode by using the FORK command and is returned to the global mode by the JOIN command. When the JOIN is issued, the program will not continue operation until all subjects have completed the previous program segment. A subject program can reference only global variables in the global mode, while either global or local variables can be accessed in the local mode.

The Operating System

A discussion of the PL/E language is incomplete without a description of the underlying operating system. To a large extent the EMPP system can be described as an operating system without an operating system. The only required element of the resident system is a 100 instruction kernel, the scheduler. The scheduler consists of basically two submodules; the interrupt service subroutine (ISS) and the background scheduler or interpreter.

The interrupt service subroutine assigns priorities to the various I/O devices and on detection of interrupts passes control to the appropriate device handlers. If after processing an interrupt ISS determines that the system was previously idle, control is passed to the background scheduler.

The basic scheduling algorithm of the system is first come first served. A subject retains control of the machine until issuing a request for an I/O operation or a delay. When the operation is complete, the subject is appended to the end of an execution queue.

This very simple scheduling algorithm is sufficient for a large number of real-time systems. The only constraint is that time critical operations must be processed with a priority greater than that of normal programs. Thus for example, timing measurements for the consoles are performed as soon as the appropriate interrupts are recognized. There may be a delay in returning to the subject's program, but the recording of data is well within the timing accuracy of the system (1 msec).

Execution of Experiment Macros

Figure 3 is a block diagram of the EMPP operating system. Subject programs

----

Insert Figure 3 about here

----

and experiment macros communicate with each other and the scheduler through a series of doubly linked lists or queues. The scheduler operates on a single queue, the execution list. Whenever a subject is suspended, it is placed on a device queue and when ready for reexecution is appended to the execution list.

With non input/output commands the scheduler passes control directly to the experiment macros. The experiment macro performs the desired operation and returns control directly to the scheduler. The scheduler then selects the appropriate instruction for the subject and continues.

In the case of input/output operations the scheduler first branches to a device monitor that determines if the requested device is free. If the device is free, the input/output operation is started and the subject is placed at the head of the device holding queue. When the I/O operation is completed, the subject is removed from the device queue and appended to the end of the execution list. If the requested device is busy, the subject is appended to the device holding queue. When the device becomes free, waiting subjects are processed by the device monitor on a first come first served basis.

Compiled Code - The Meta-language

The code generated by the compiler is in a three address Polish prefix format. For example, I = x+Y would be compiled as

```
ADD
@x
@Y
@I
```

were ADD corresponds to an integer constant pointing to an interpreter branch table and X,Y and I are pointers to the dope vectors of the appropriate variables. The calculation of variable addresses is performed by a single system subroutine (GETARG) which examines the dope vector to determine the data address, verifies that the address is valid and returns the address of the argument.

This very simple organization makes it easy to add either pseudo instructions (commands) or new data types to the PL/E language without having to modify existing software. New commands are added to PL/E simply by adding the command name and information about its parameters to a system data file read by the compiler at run time. New data types or structures can also be added to the system with relative ease. For example, structured data records (ala COBOL or PL/I) could be added to PL/E simply by modifying GETARG and adding a new data type definition command to the compiler.

Evaluation of the System

The EMPP system stands up quite well to its initial design goals. Experiment programming is now fairly simple for persons unfamiliar with computers. The primary benefit of the system is that programming time can be expressed in terms of hours rather than weeks.

The PL/E language has proved to be quite powerful allowing the development of several fairly complex experiments that had previously been avoided due to programming difficulty. PL/E greatly simplifies the development of experimental programs by offering psychologists a powerful tool for expressing experimental algorithms and for documenting experimental procedures.

Methodology of Design

The successful completion of the EMPP system is a direct result of the methodology used in its development. The technique used can best be described as an adaption of the heuristic iterative enhancement method (6).

Dijkstra and others (7, 8) have discussed the benefits of a well structured top-down approach in producing easily modifiable software in a relatively short period of time. In real-time systems such as EMPP, however, this technique is difficult to apply. To construct a system using a pure top-down approach requires a complete understanding of both the problem and at least at the top level its solution. In many laboratory systems it is not possible to completely specify the problem. A research environment by its very nature requires a flexible solution.

Iterative enhancement (6) is a heuristic technique related to the top-down algorithm that involves selecting a "doable" subset of a loosely defined problem, implementing the subset in a top-down fashion and then refining the partial solution. A larger subset is then selected and the process is continued until a final solution is obtained. If at any point in the iterative process problems result, it may be necessary to backtrack and redesign one of the previous solutions.

In a system such as EMPP the difficulty is that this final step is never really made. As the powers of the subsets increase, the demands on the system increase. As refinements to the solution are made, new problems come up and the system constantly evolves. That is not only the solution, but also the problem evolves with experience.

Implementation of the System

Figure 4 is a table of the steps taken and the time required to implement the various phases of the project. It required approximately one man years

Insert Figure 4 about here

to complete the system.

In addition to the previously mentioned environmental considerations, there were several practical problems that had to be resolved during the implementation of the system. First, several new hardware devices were going to be added to the system within six months. In particular, a new disk and scope controller were going to be designed, however, neither had been selected. Second, minimal computer time was available for debugging because the laboratory was in the midst of an experiment marathon, a large set of production experiments. Third, there was a need to get a least a skeletal operating system running within three months to enable the inclusion of several new experiments in the marathon.

The Global Pass

The first or global pass at the system involved an overview of the requirements of the laboratory from a functional standpoint and a very rough attempt at specifying a language that would be applicable. An attempt was made to determine what types of experiments would be performed in the laboratory and what tools would be needed to implement them. During this overview of the system it became obvious that it was impossible to completely specify all experiments to be run on the system and that EMPP would have to be modularized to the extent that any portion of the system could be modified without effecting the remainder. Potential modifications could range from adding a new input/output device to changing the scheduling algorithm.

Implementing a Preliminary Operating System

The second pass involved the implementation of a skeletal operating system. During this phase, the majority of the system parameters were specified in detail. The decision was made to implement the tentative language with the aid of an interpretive meta-language. Aside from the potential benefits of portability, computer scientists have a tendency to dismiss interpretive approaches due to slow speeds of execution and increased memory requirements. In many on-line applications, however, an interpretive approach is superior to machine language. In systems where there are multiple copies of the same program, compiled code can be memory inefficient. In addition, given the limited instruction sets of most minicomputers, it is difficult to generate efficient compiled code. Even the standard language compilers supplied by most minicomputer vendors rely heavily on interpretive techniques for implementation.

The first come first served scheduling algorithm was selected as the easiest to implement and the easiest to test. Experience with the system has shown that it is sufficiently powerful for our needs.

All systems data formats (stacks, queues, dope vectors, etc.) were selected and a series of system subroutines were written to operate on them. After coding the scheduler and testing a few simple experiment macros, the requirements of formatted I/O were laid out and routines were written to interface the operating system with existing I/O drivers.

Production Environment Testing

The third phase of the system development was to test the skeletal operating system in a production environment. Several experimental programs were hand coded in the meta-language and run as part of the experiment marathon. No great problems were detected with the new system, however, as experiments were being coded, a number of new experiment macros were added to the instruction set of the pseudo machine.

Compiler Implementation

The fourth phase of system development was the implementation of a simple compiler. Although somewhat primitive in that it just recognized command statements, the compiler was sufficiently powerful to allow psychologists to develop their own experiments. A polish prefix notation was selected as the easiest method for developing a compiler in a short period of time. Although the resulting programs were not strictly readable in the natural language sense, they were a great improvement over assembly language. There were a few simple rules to learn, but once these rules were mastered programming was relatively simple. The major difficulty with the first version of the language was its unwieldiness for arithmetic calculations and its lack of conditional expressions.

At about the same time that the compiler was being written some new hardware arrived in the laboratory. The disk and new scope controllers were interfaced to the existing software. Neither one presented any problem even though the new scopes were significantly different than the previous variety.

System Refinement

After putting the system in a production environment, users had the opportunity to gain a good feel for the advantages and the weaknesses of the EMPP system. During the fifth phase user feedback lead to several minor changes to the compiler to improve program readability and to decrease the difficulty of writing programs. Other systems refinements were adding a graphics package and improving compile and run time diagnostics (e.g. describing an error in detail rather than saying the disk just died or there is a format error in statement x). It was during this phase that the syntax and semantics of the present version of PL/E were defined.

The Completed System

At present the fifth phase has been completed and development is entering the sixth or final phase. (Although in fact there may be more modifications to the system in the sixth phase than in the previous three phases.) The language has been extended to support algebraic statements, conditional statements and iterative loop control. The compiler has been written and tested and is now being used in a production environment. Although most users were comfortable with the previous version of the compiler, users have found the new version of the language to be significantly easier to program and to debug than earlier versions.

Conclusion

The successful completion of the EMPP system has demonstrated the utility of the iterative enhancement technique in the design of a loosely defined on-line system. By successively selecting larger subsets of the initial problem it has been possible to generate a powerful general purpose system in a relatively short period of time. The power of the EMPP system can be attributed to the simplicity of its underlying operating system. Although the PL/E language is fairly complex, the underlying operating system has been kept as simple as possible to speed development and aid in debugging.

Similar techniques can probably be used to develop on-line systems for other application areas. Although PL/E as implemented in the EMPP system has a fairly extensive repertoire of commands, it would certainly be possible to add or delete commands for another application area. The same is true for the system scheduling facilities. The same simple scheduling philosophy could be used to implement systems with multiple priority levels of execution or in systems with only a single subject.

Perhaps the best measure of success of any on-line system is the acceptance of the system by its users. From this standpoint the EMPP system has been successful. Several psychologists with no previous computer training are now actively developing their own experiments. The users of the system can now concentrate on the psychological aspects of experiments and work with the computer system as a tool rather than as an advisary.

References

1. D. W. Zobrist, et al., "Software standards and CAMAC... A realtime demonstration," Instrumentation Technology, pp. 33-38, March 1975.
2. G. F. Pitz, "Building a programming language for a small computer: Reinventing the wheel," Behavior Research Methods & Instrumentation, Vol. 7,No. 1, pp. 42-46, January 1975.
3. D. J. Getty, "The PEPL system for control of experiments by a PDP-8 computer," Behavior Research Methods & Instrumentation, Vol. 7,No. 2, pp. 131-136, February 1975.
4. A. L. Pinkus & L. W. Gregg, "APCOL: A programming system for computer-controlled psychology laboratories," Behavior Research Methods & Instrumentation, Vol. 5, No. 2, pp. 65-172, March, 1973.
5. P. G. Polson, "SCAT: Design Criteria and software, "Behavior Research Methods & Instrumentation, Vol. 5, No. 2, pp. 241-244, March 1973.
6. J. E. Basili & A. J. Turner, "Iterative enhancement: A practical technique for software development," IEEE Transactions on Software Engineering, Vol. 1 SE1, No. 4, pp. 390-396, December 1975.
7. O. J. Dahl, E. W. Dijkstra, & C. A. R. Hoare, Structured programming, London: Academic, 1972.
8. N. Wirth, "Program development by stepwise refinement," Commun. As. Comput. Mach., Vol. 14, pp. 221-227, April 1971.

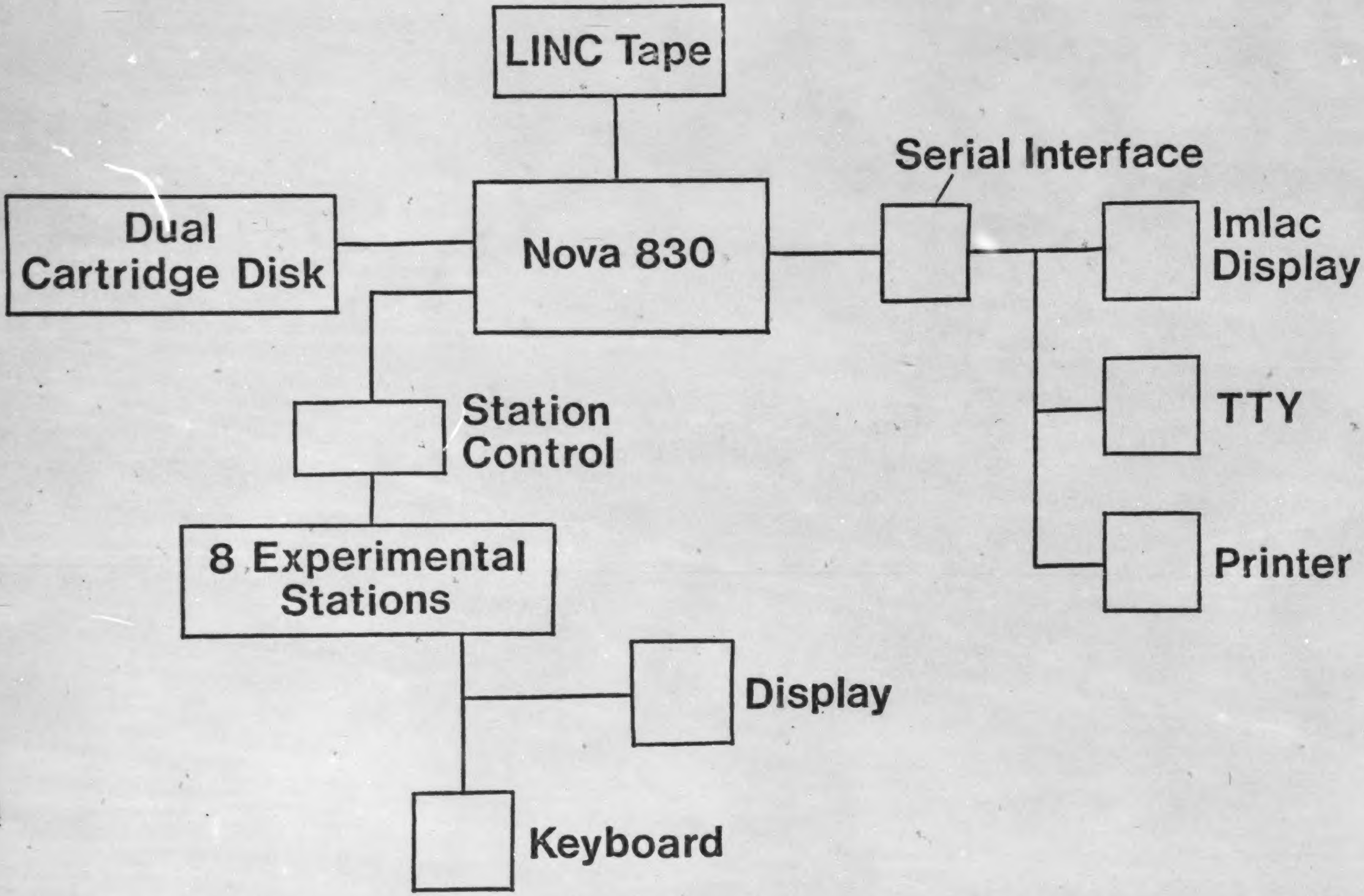
Figure Captions

Figure 1. EMPP System Hardware

Figure 2. EMPP System Software

Figure 3. The EMPP Operating System

Figure 4. Implementation of the EMPP System



LINC Tape

Dual Cartridge Disk

Nova 830

Serial Interface

Imlac Display

TTY

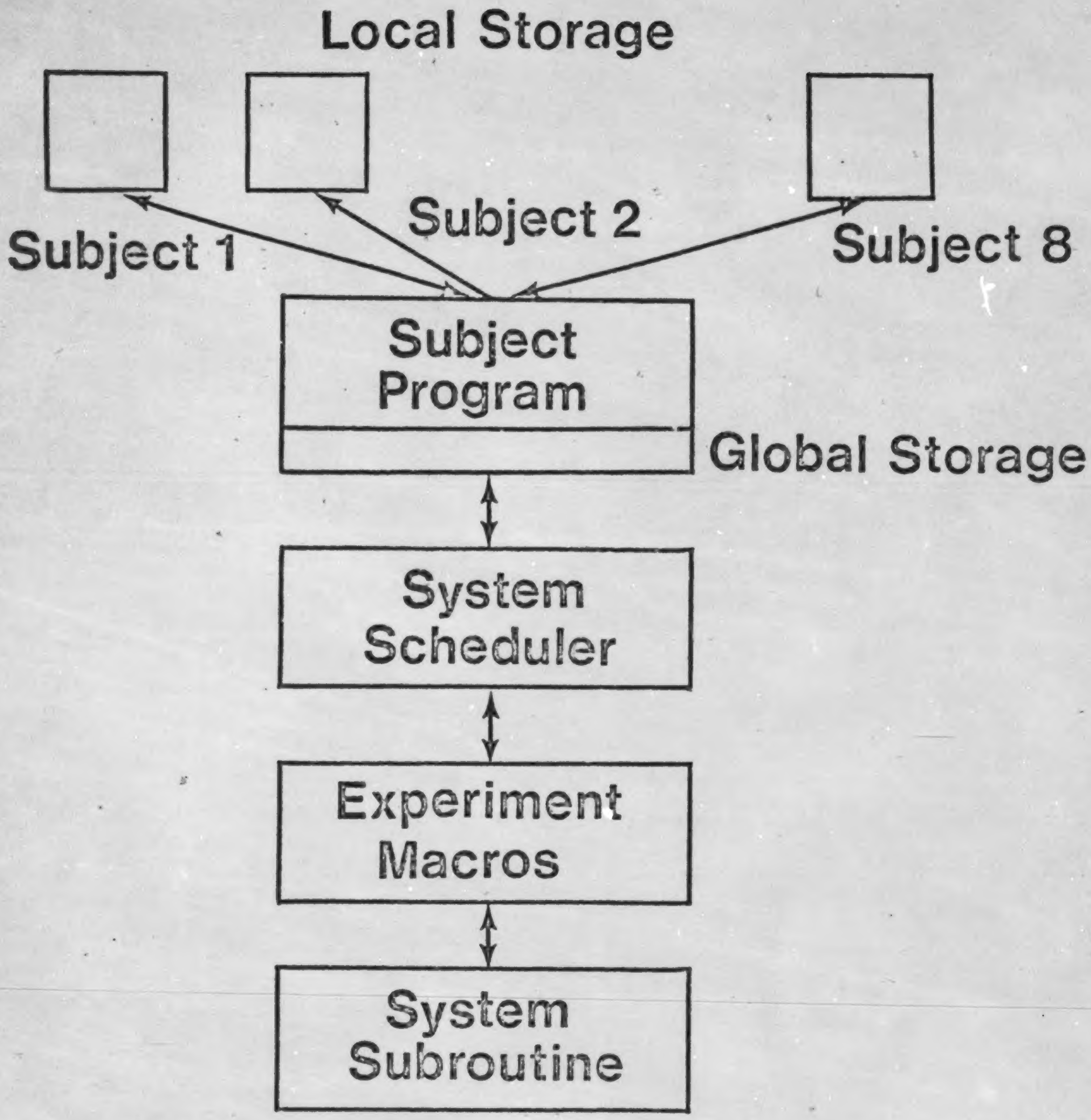
Printer

Station Control

8 Experimental Stations

Display

Keyboard


Local Storage
Subject 1
Subject 2
Subject 8
Subject Program
Global Storage
System Scheduler
Experiment Macros
System Subroutine

Verbal Input for the EMPP System

Kenneth J. Burkhardt, Jr.

The EMPP system is an integrated hardware and software system designed to simplify the development of on-line psychological experiments. The system is presently capable of communicating with eight subjects and an experimenter by means of a set of keyboard consoles and CRT display terminals. Experimental stimuli are presented at the display terminals and subjects respond via specially designed control consoles. The power and flexibility of the system could be greatly increased if subjects could respond verbally rather than having to physically press buttons on the consoles.

The purpose of this paper is to describe the problems encountered in attempting to design an on-line speech input terminal for the EMPP system. The paper consists of three sections. The first summarizes the desired characteristics of the prospective terminals. The second outlines the state of the art in speech recognition and the third describes a potential hardware organization for a speech input system.

Characteristics of the Input Terminals

The term speech recognition has been used to describe a broad range of problems varying in complexity from simple yes/no type applications to the presently unsolvable problem of spoken English. One of the difficulties in dealing with speech recognition is that there are no firm procedures for defining a speech recognition problem and its solution. Even the presently accepted performance measures of speech recognition systems in terms of number of words and speakers that the system can handle are often meaningless (depending upon the phonetical differences in the recognized vocabularies). Accuracy is also a weak measure since a 90% accurate system in real-time may be "better" than a 98% system in 100 times real time (Reddy, 1972). The approach most researchers in this area have taken is to select a subset of the general speech recognition problem and then attempt to construct a system that can be modified in an iterative fashion as the researcher approaches the desired goal.

One of the problems in attempting to design speech input terminals for the EMPP system is that we are not in the business of research in speech or pattern recognition. Our interests are in psychology and therefore we are interested in speech input terminals as a tool rather than an end. Thus we have neither the resources nor the desire to conduct research in speech recognition, but rather would prefer to buy our terminals "off the shelf." Unfortunately, this is not feasible at the present time. It seems probable, however, that within the next decade there will be a wide variety of commercially available speech input terminals.

What is required for the EMPP system is a terminal that is functionally equivalent to the existing push button consoles. That is the speech terminals should be capable of recognizing responses chosen from a small set of potential answers. Response must be in real time with an accuracy of response of 100% and a resolution of 1 msec (from start of the utterance). Furthermore the system must be capable of supporting eight stations where operations at each terminal are uneffected by operations at other terminals.

In terms of presently available speech systems this is a fairly complex problem. There are, however, several simplifying assumptions that can be made. For example, the set of responses will never contain more than ten elements. Furthermore, constraints can be placed on the operation of the system, such that each subject can be required to "teach" the system his response set, i.e., Speech input terminals do not have to be totally independent in that a particular speaker can be assigned to a specific terminal. (This eliminates many potential speech recognition problems caused by differences in individual speech patterns.)

From the standpoint of implementing the proposed system, the most serious technological problems center around the requirements for 100% accuracy and for real-time response from multiple subjects. Most existing systems, even those centered around large computers, cannot meet these requirements. Before discussing a potential solution to these problems, it will be useful to review the present status of research in automatic speech recognition.

Current Research in Automatic Speech Recognition

In spite of more than two decades of research in the field of automatic speech recognition there has been limited success. This slow development cycle can be attributed primarily to two factors: inadequate technology and incorrect models of speech. The result has been that although a large number of systems have been designed on paper, there are actually relatively few working speech input systems (Reddy, 1972).

Figure 1 summarizes the result of some of the more recent attempts at speech recognition. Londgren (1965) and Hyde (1965) provide excellent surveys of the state of the art prior to 1968. Present systems claim to achieve from 60% to 99% accuracy of recognition. However, as mentioned earlier it is somewhat difficult to quantitatively compare the relative merits of these systems. Although accuracy is a convenient measure, one must also consider recognition time, vocabulary size, complexity of the words, and hardware needed to implement the system.

One difficulty in applying present speech research to the development of a speech input terminal for experimental psychology is that the primary concern of researchers in the speech recognition field is with continuous speech and relatively large vocabularies. Thus the equipment used in many speech projects is totally incompatable with equipment that could be supported by our system. For example, the CMU speech project uses a dedicated PDP10 connected to various special purpose hardware devices to recognize speech from a single subject. Strassbourger's speech recognition project at the University of Washington required a dedicated Sigma 5 and was still not capable of processing input from a single subject in real time.

Speech Recognition for Limited Vocabularies

Attempts to recognize limited vocabularies using "feature extraction" classification methods of classical pattern matching have met with some success (Talbert, et al., 1963). The primary problem in dealing with small vocabularies is resolving just what parameters are required and what types of analysis should be used to extract these parameters.

Speech signal measurements are generally digitized readings taken in the time domain and then processed directly or converted to the frequency domain by means of Fourier transformations. Other techniques are integration in the Humph/Hiss domain (Hill, 1971), counting zero crossings, cepstrum (Strassbourger, 1972) analysis, integration within frequency bands, etc. The primary difficulty in gathering speech data is the high bandwidth (10=20 KHz) required by unfiltered speech data in the time domain. A secondary difficulty is encountered when data is to be transformed to the frequency domain. Unless the user can afford a special purpose array processor, Fourier transforms can require more time for calculation than is available to process data on the fly. For example, the PDP11/20 requires 1 sec to perform a simple FFT on a 1000 point sample. Thus the maximum data rate that can be processed is 1 KHz. This is obviously too slow for realtime speech processing.

Terminal Design

Although there are presently no commercially available systems that fit the needs of the Psychology Department, I believe that within the next decade terminals will either be marketed or components will be available to justify our installing a system. With this in mind the following is a brief description of the architecture of a potential speech recognition system.

Figure 2 is a block diagram of the proposed system. The system will consist of nine computers; one master control unit (MCU) and eight satellite processors. The master control unit will oversee the operation of the eight satellites and will control communications with the host computer (the NOVA). The MCU will also contain a secondary storage device (e.g., a disk) for maintaining a backup copy of satellite processor control programs and for maintaining directories of vital speech parameters for each subject in the subject pool. When the central computer assigns a subject to a processor (experimental station), the master control unit will send the subject's vital parameters to the appropriate satellite which will then be prepared to collect and process the on-line speech data.

The actual configuration of each satellite processor will be a function of both method of analysis selected and the state of the art of computer components at the time the system is implemented. With the ever increasing speed and decreasing cost of arithmetic processing units it is possible that the satellites will be completely digital. In this case data would probably be collected in the time domain and then Fourier transformed to the frequency domain. The satellites processors, however, will probably be part digital and part analog in which case transformations may be partly handled by hardware filters. The speech research system at Carnegie Mellon University uses a similar technique. In this system a number of hardware filters are used to extract sample data in set frequency ranges and count zero crossing automatically. At present, however, this technique is too expensive in that it requires a dedicated PDP10 for a single subject.

Another potential technique might be to modify a commercial speech input terminal such as the VDETS 1000 series advertised by Scope Electronics of Reston, Virginia. These terminals probably use a combined software/hardware approach to give close to real-time response for input. (I have been unable to obtain detailed information on the system, however, and thus suspect that it is only in the prototype stage and will probably be fairly expensive when introduced.) In this situation the only problem on system design would be to develop the proper protocol for master to satellite communications.

References

Hill, D. R. Man-machine interaction using speech. In Advances in Computers, 11, New York: Academic Press, 1971, 165-230.

Hughes, G. W. & Hemdal, J. F. Some experiments in spoken word recognition, IBM J. of R. and D., 10, 1, 65-79 (1966).

Hyde, S. R. Automatic speech recognition: Literature survey and discussion. RDR No. 45, Post Office Research Department, Dollis Hill, London N.W. 2 (1968).

Londgren, N. Machine recognition of human language. IEEE Spectrum 2, Nos. 3 and 4 (1965).

Newell, A. et al. Final report of a study group on speech understanding systems. Computer Science Department, Carnegie-Mellon University, Pittsburgh, 1971.

Reddy, R. et al. Working papers in speech recognition. Computer Science Department, Carnegie-Mellon University, Pittsburgh, 1972.

Strassbourger, E. Recognizing a continuous spoken language using analytic cepstrum encoding. Ph.D. Thesis, Computer Science Department, University of Washington, 1972.

Talbert, L. R. et al. A real-time adaptive speech recognition system. Technical Report No. 670-1 (ASD-TDR-63-660), Stanford Electronics Laboratory, 1963.

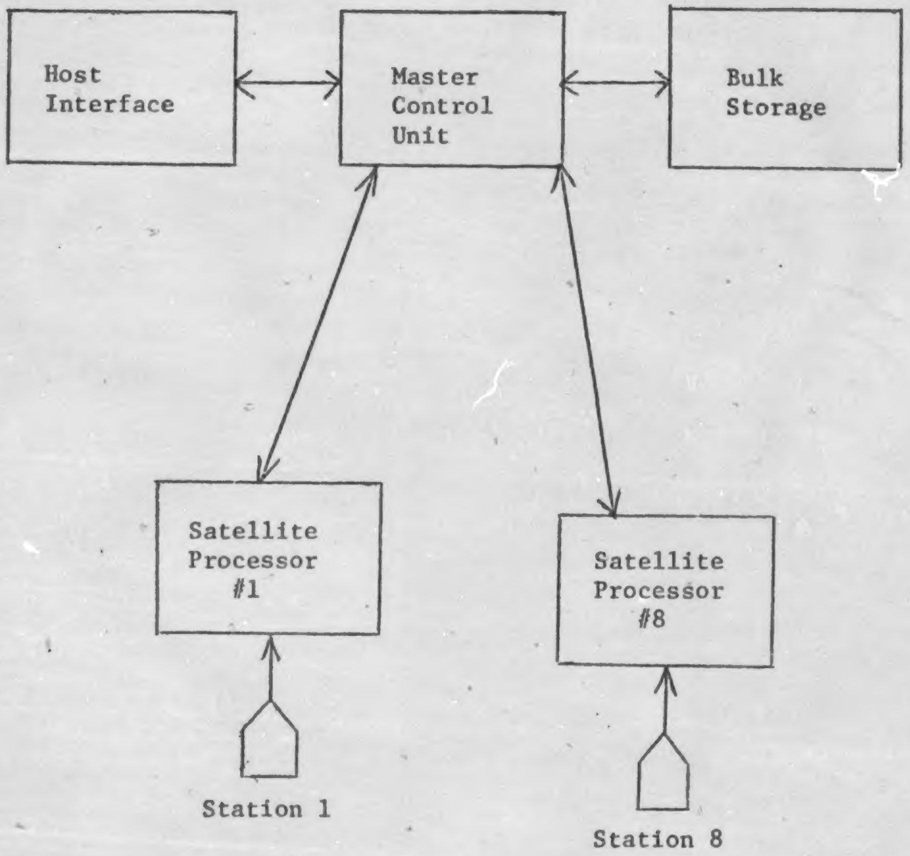
Vicens, P. J. Aspects of speech recognition by computer. Ph.D. Thesis, Computer Science Department, Report No. 127, Stanford University, California, 1969.

Figure 1

System	Year	Dominant Domain	Vocabulary Size (in words)	Accuracy
Vicens	1969	Time	561	92%
Hill	1969	Humph/Hiss	16	78%
Elder	1970	Frequency	37	96%
Clark	1970	Frequency	10	71-96%
Glen	1971	Frequency	24	Not reported
Clapper	1971	Frequency	16	87-96%
Pols	1971	Frequency	20	70-95%
Herscher	1972	Frequency	21	99%

Some speech recognition systems
(See Strassbourger (1972) for references)

Figure 2


Host Interface
Master Control Unit
Bulk Storage
Satellite Processor #1
Satellite Processor #8
Station 1
Station 8

Hardware configuration for a multi-speaker speech recognition system.

I. Introduction

Overview of the proposal. This is a proposal for a continuation of our study of individual differences in cognition. The long-term goal of our project is to establish a theory of intelligence which is derived from a general theory of cognitive processes. In this section we expand upon the reasons for such a theory, and consider the form it might take. The next section contains a progress report on work during the past four (of five) years supported by NIMH grant MH21795. The progress report section is itself divided into three subsections; a summary of the state of knowledge at the beginning of the grant period, a description of our own work, and an assessment of revisions in our knowledge based upon the work by ourselves and others. The third major section of the grant consists of proposals for future research. This contains a series of research questions generated by the revised model, together with experimental studies designed to answer them.

Rationale. Intelligence, in some gross sense, is so ubiquitous a human trait that when a psychologist develops a personality test, s/he often tries to show that the test does correlate with intelligence! We shall not bother to document the impressive array of statistics describing the correlations which exist between intelligence test scores and various measures of psychological and sociological interest. We believe that virtually every psychologist accepts the existence of some form of "cognitive ability," acknowledges that it differs between individuals, and recognizes that the difference is important. Psychologists differ drastically in their assumptions about what cognitive ability is, and why it shows such interesting variation. Those who disbelieve this are invited to consider the long "nature-nurture" controversy.

De facto, the most influential theories of intelligence are theories of what we shall call psychometric intelligence, i.e., cognitive ability as demonstrated on typical "IQ" and aptitude tests. These tests have a long and well known history, stretching from the immensely influential work of Binet to the current, highly sophisticated analyses of Guilford, Cattell, and other modern psychometricians. Psychometric theories of intelligence have two important characteristics. They are closely tied to a mathematical model which locates individuals in a factor space defined by their relative abilities to deal with an assortment of problems chosen to display different aspects of intelligence. (The single IQ score is, in this case, the mathematically degenerate situation in which the factor space is one-dimensional.) In spite of their elegant mathematical structure, psychometric theories are basically concise descriptions of data. Psychometric measures indicate the relative placement of individuals with respect to each other without any commitment to a process model of cognition. Cognitive processes may be inferred from psychometric scores (e.g., in Guilford's structure of intellect model (Guilford, 1967; Guilford & Huepfner, 1971) or in Cattel's (1971) distinction between fluid and crystallized intelligence), but process identification is based on an act of inductive inference on the part of the psychologist. Naturally, there is nothing wrong with this, but it is not the only way that one might study individual differences in cognition.

We wished to consider an alternative approach, based on the deductive paradigm which characterizes virtually all of experimental psychology. In the deductive approach, hypothetical models of cognitive processes are established, and the parameters of these models are evaluated under varying conditions. The deductive approach has the advantage of being concerned with the dynamics of cognition per se. The measures taken fit into a model of the information processing that is going on when one thinks. To the theoretician a measure of speed of access to verbal information in long term memory is a more sensible statement about ability than a measure of relative standing on a test of comprehension of sentences.

Even acknowledging this advantage, it is not obvious that the deductive approach is to be preferred in the study of intelligence. The psychometric approach to intelligence is bound to generate a useful data summary, whereas the deductive approach may not, if the theory behind it is wrong. In the abstract, a deductive theory of intelligence should wait for a complete theory of the process of cognition. A partial theory of intelligence depends on the partial truth of its cognitive model. This is the problem which defeated Galton's efforts to establish a theory of individual differences, and which hampered a number of efforts to tie intelligence to learning theory in the 1930s and 1940s. Today's models of cognitive processes are far more sophisticated than the models those investigators used. Whether they are adequate models on which to base a theory of individual differences remains to be established.

The psychometric approach also works in a pragmatic sense. Suppose one's purpose in having an intelligence test is selection of personnel for further training or for special care. This, after all, was Binet's motivation. The validity of the measure is then determined by its correlation with appropriate criterion variables, quite apart from any theoretical rationale. Furthermore, since we are talking about a personnel selection procedure, questions of cost effectiveness enter. Latency and physiological measures cost much more than paper and pencil tests. Thus it is doubtful that the deductive approach to intelligence will generate tests that can compete in the applied market. What it can do is generate tests which can compete in the "market" for tests intended to aid in diagnosis of individual cases. If we want to know what a person's cognitive power is, relative to some criterion group, a psychometric test is appropriate. However, theoretically based cognitive measures have the capability of identifying specific deficits and abilities which contribute to psychometric performance.

The usefulness of theoretical tests is even greater whenever a general scientific question about individual differences is to be addressed. One would think it obvious that concept A cannot be scientifically related to concept B by observation of covariation between measurements A' and B', respectively, unless the A-A' and B-B' connections are clear. Surprisingly, this has not always been the case in studies of intelligence. The best example is the prolonged debate concerning genetic influences on intelligence. Some of the literature reads as if the key question is whether "potential for scoring at level X on test Y" is inherited or acquired. It could not possibly be either, alone. Physiological structures are inherited, and these mechanisms are reflected in process measurements, e.g., in the scanning of short term memory. Test scores are acquired by exercising one's mental mechanisms through the use of specific mental algorithms ("strategies") which operate on stored information. The mental structures must be produced by physiological variables (including genetic ones), and the stored information must be learned. The algorithms vary...in some cases the mental structures permit only one algorithm and in other cases the algorithms may be learned. Within a theoretical model of intelligence we have to understand the interaction between structures, algorithms, and information. It is our contention that this can only be done by relying upon theoretically based measurements.

II. Progress report

Cognitive theory circa 1970. As indicated above, our research program is inevitably a series of jumps. We begin with a cognitive theory, and apply it to individual differences. Our research will, naturally, cause us to modify the theory. At the same time, however, results obtained by others, who are studying cognition but not individual differences, will also force modifications in the underlying cognitive theory. The combination of new data on both individual differences and cognition will force a rethinking of the original model, which will then lead to a new derived model of individual differences. We shall begin with a picture of the cognitive theory which we held at the beginning of our work. We shall then present our own results, and comment on the related research done by others. Finally, we close the progress report by stating our revised view of the general cognitive models most relevant to individual differences. This will provide a basis for the discussion of our proposals for new research.

Cognitive theory circa 1970 could be characterized by three themes relevant to individual differences. The first was an emphasis on temporal stages of processing. Atkinson and Shiffrin's (1968) very well known structural model and our own ideas of the mechanisms of information processing (Hunt, 1971) postulate a brief "pre-perceptual" memory system, followed by a short term, active memory, and then a much larger, almost permanent long term memory. (We point out that the "long term memory" of the Atkinson and Shiffrin model temporally corresponds to the "intermediate term memory" stage of the Hunt model.) Research interest focused on the existence of bottlenecks in the information processing system. Short term memory was seen as the greatest problem, since information had to be consolidated and rearranged in a store which was believed to be very small. The entire structure was thought to be "programmable" in the sense that different forms of information could be consolidated by selection of the information to be retained in short term memory. In a theoretical paper, we demonstrated how the rather primitive system just outlined could be "programmed" to achieve quite complex problem solving by a continuous interchange of information in short term memory (Hunt & Poltrock, 1974). We used the same information handling formats used by Newell and Simon (1972) to create simulations of very complex problem solving, so we felt that we had established a conceptual tie between the "elementary" algorithms used to exercise mental structures on small problems and more elaborate algorithms used to solve such complex problems as chess and cryptarithmetic.

There are several ways to interpret 'stages' in a stage of processing model, and there has been considerable theoretical controversy over the proper one. For our purposes in the analysis of individual differences, though, the question was not so much what the model can encompass as what it stresses. Three classes of parameters seemed important to measure:

1) The size of critical buffers, and especially the short term memory buffer.

2) Rate of transfer of information from one buffer to another.

3) Measures of the accessibility of information within each memory store.

These parameters all reflect the performance of general memory structures. Some or all of the stages of memory might be sensory specific. Perhaps we should speak of different abilities for dealing with visual, acoustic, tactile, and even olfactory memory. Even more important, individuals might differ in the algorithms which they use to solve the same problem. We would expect this to be the case if people vary in their structural capacities...each person ought to choose those algorithms which depend most heavily upon that person's strengths. We can distinguish two possibilities. It might be that people vary little in structural capacity. If so one would expect to find that some people develop efficient and others inefficient problem solving strategies, but that the goodness of a strategy can be assessed independent of the person using it. On the other hand, if people vary in their structural capacities, the goodness of a strategy will depend upon the structural parameters of the person using it.

Individual differences must always be defined with respect to a particular set of tasks. The simple finding that "there are individual differences," without assessing their importance, is not interesting. But importance for what? Suppose a parameter is found to vary over individuals. This is interesting only if it either (a) varies over such a range that, within the population, the success or failure of some individuals at some task is virtually assured by the parameter value, or (b) if the variation is such that some individuals are forced to adopt one strategy on a task, and some individuals are forced to adopt another. Whether this is true or not depends on the task. Ideally there would be a theory of problem solving which would tell us which tasks are basic. In practice, there is not. In fact, reviewers (e.g., Johnson, 1972) have pointed out that the psychological study of problem solving has been marked by arbitrary choices of problems and, as a result, non-generalizable findings.

Given this situation, we made a strategic decision to use a subject's performance on a standard verbal intelligence test as a criterion for interesting differences. There were two reasons for this choice. The first was that verbal intelligence is the most ubiquitous factor in intelligence test research. It also has considerable practical value. In spite of the bad press which it has received recently, the verbal intelligence test score is the best single predictor of general academic performance (Lanvin, 1965). Since "general academic performance" samples a wide variety of problem solving behavior, the verbal intelligence test is probably the best single, easily available measure of problem solving skills.

The reasoning outlined above dictated our decision to study particular variables. In addition, we made a decision to concentrate publication in a small number of fairly lengthy papers, rather than reporting isolated studies. We chose this route because the principal investigators feel strongly that the history of intelligence as a psychological concept, and particularly the recent bitter debate over genetic and racial correlates of intelligence, has suffered from the failure of psychometricians and cognitive psychologists to consider each others' point of view. We hoped that by publishing a small number of substantial papers we would do more to correct this situation than we could by publishing the same amount of research in smaller papers in various specialized journals.

Progress Report: Work Completed and Underway

Overview. Our work is best considered in four "streams." The first, and by far the largest, consists of studies of information processing correlates of verbal intelligence within the university or high school population. We also report some studies of unusually gifted individuals within these populations. The second stream consists of studies relating psychometrically defined verbal intelligence test scores to skills involved in the perception and manipulation of the spoken language. In spite of the name, the verbal intelligence tests which we have used do not require any reaction to the spoken language. On the other hand, most psychologists feel that some sort of internal language plays a major role in problem solving. Thus studies of language manipulation in persons with varying verbal intelligence test scores may directly address the question of why some people are, generally, good problem solvers.

Our third stream of studies expands the first two by moving to a rather different level of information processing. As will be seen, the experiments in the first and second streams are concerned largely with the information processing involved in the execution of presumed basic processes in whatever kind of computer man is (Hunt, 1971). These experiments involve situations in which people must access overlearned information, scan short term memory, and do other similar tasks in very simple situations. In the third stream of experiments we look at the relation between intelligence and strategies for problem solving. This work is in a less completed status because initiation of the experiments often depended upon the completion of the more molecular studies in the first and second stream. Our proposals for future research deal with this topic at length.

The final stream consists of experiments applying the information processing approach to psychological measurement in the fields of mental retardation and evaluation of the effects of drugs upon cognition. These experiments were conducted both for their own inherent interest and because they provide extreme test cases for the theoretical approach. The results obtained have suggested some questions for future research. In addition to these four streams, we shall also report some research related to applied problems in psychometrics.

Studies of verbal intelligence. Our most frequently used paradigm contrasts the performance on experimental tasks of people who have high or low verbal intelligence test scores, so we shall routinely speak of "high verbals" and "low verbals." Unless noted otherwise, these terms refer to people in the top and bottom quartile of verbal intelligence test scores within the population of University of Washington students. The "high" group would be high in almost any population, while the "low" group is probably about representative of the *middle* third of the general population.

The term "verbal intelligence" is likely to connote more than it means. In spite of the name of the test, one never has to speak or respond to a spoken message in order to earn a high verbal intelligence score. Most tests require responses only to the written language. Although in some cases the tests do have a speed component, speed is measured in terms of minutes rather than the millisecond timing required to measure cognitive processes. Furthermore, it is clear on its face that in order to obtain a high verbal intelligence test score one must have culture-specific information about standard American English and a good grasp of the customs of Middle Class America. No one doubts that there is a learned component to test taking. The interesting theoretical question is "Does test performance reflect only the amount of exposure to the necessary information, or does it reflect information processing parameters which do not depend on knowledge?"[1]

Experimental studies I. *Memory in high and low verbal subjects.*

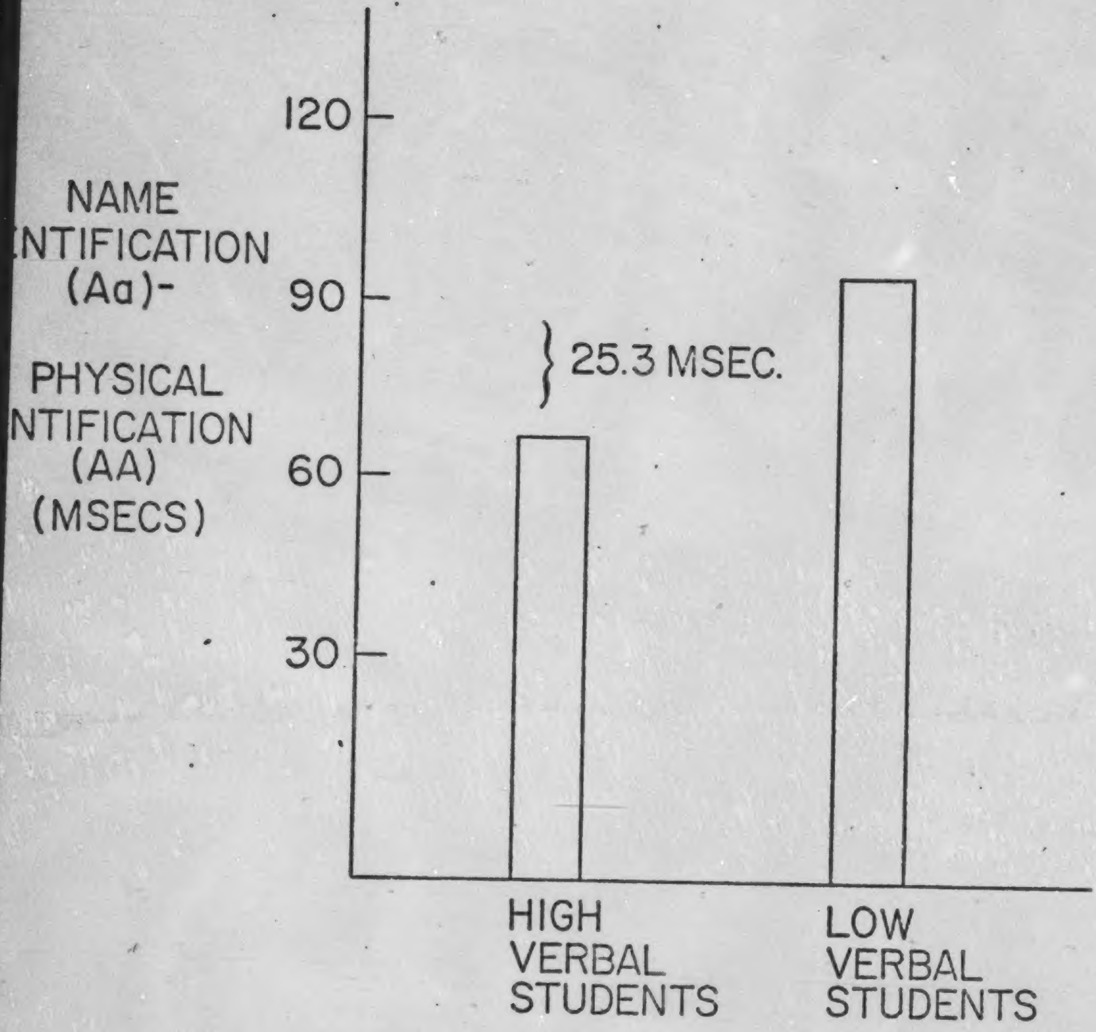
This section summarizes the experiments reported in Hunt, Frost, and Lunneborg (1973); Hunt, Lunneborg, and Lewis (1975); Hunt (1976), and Hunt (in press, a). Unless otherwise noted, these studies used a design based on the 'high verbal - low verbal' distinction.

We first consider tasks which measure access to highly overlearned information in memory. Most of our everyday memory tasks, such as recognizing that a particular pattern of lines and curves means STOP, depend on this. Posner and Snyder (1975) have referred to such access as *obligatory coding*. Posner and his colleagues (Posner & Mitchell, 1967; Posner *et al.*, 1969) measured the speed of obligatory coding in the "Name Identification-Physical Identification" (NI-PI) task. The subject is required to identify two letter patterns as *naming* either the same or a different letter. Thus the response to "AA" and "Aa" is "same," while the response to "AB" is "different." Responding to an "AA" pair, however, does not require that one ascertain the name of the patterns, for the names of two identical patterns must be identical. Name determination is required to ascertain that "A" and "a" refer to the same letter. Refer to these conditions as the "physical identification" and "name identification" conditions, respectively. A well known finding of Posner and his group is that name identification takes reliably longer than physical identification. The difference between the two (The "NI-PI difference") provides a measure of obligatory coding in memory which controls for motor reaction time, as the same motor response is required in each condition. In several experiments (Hunt *et al.*, 1973; 1975) we have found that there are small but consistent differences in the NI-PI measure which are associated with verbal intelligence. Some typical results are shown in Figure 1. Outside of our own laboratory, Steven Schwartz (1976) have extended this finding in an interesting way. They compared the difference between high verbal and low verbal college students in a "same sound" task using homomorphs. Here the response to DEER and DEAR would be "same." The logic is the same as in the original NI-PI task, but at a higher level of abstraction using different material. Schwartz' results were comparable to ours. The differences found are typically small in absolute terms, but they involve operations which must be done literally millions of times. Any change in the efficiency of a highly repetitive operation will be multiplied throughout an information processing system.

If *obligatory* coding is associated with verbal intelligence, it should be possible to demonstrate the relationship in a task in which the subject is not ostensibly required to retrieve a pattern name. We have attempted such a demonstration in a "memory" experiment in which the symbols to be memorized were individually meaningless, but could be juxtaposed to produce an overlearned code. The task was presented as a recall study for a list of nonsense syllables presented rapidly. In some of the lists there were pairs of syllables which, when merged, formed a word. An example is the pair "prob...lem," which merges to form "problem." This can be contrasted to the pair "lem...prob," where no such merging is possible. Now suppose the high verbal subjects are more rapid in the arousal of overlearned codes, but are *no* more advantaged with respect to their capacity to retain meaningless material. In other words, suppose that the information processing advantage of the high verbal subject is restricted to code access. In such a case, one would expect high verbal subjects to do better than low verbal subjects on lists which contain pairs of syllables that can be merged to words, but no better on lists whose members cannot be so merged. Figure 2 presents some results obtained by Hunt *et al.* (1975). The hypothesis is clearly supported. The high verbal subjects have a marked advantage in the situation in which code access helps, but have little if any advantage in dealing with memory for nonsense material. Appropriate statistical controls indicated that the difference was not due to differences in vocabularies; virtually all the university students in both groups knew the words used.

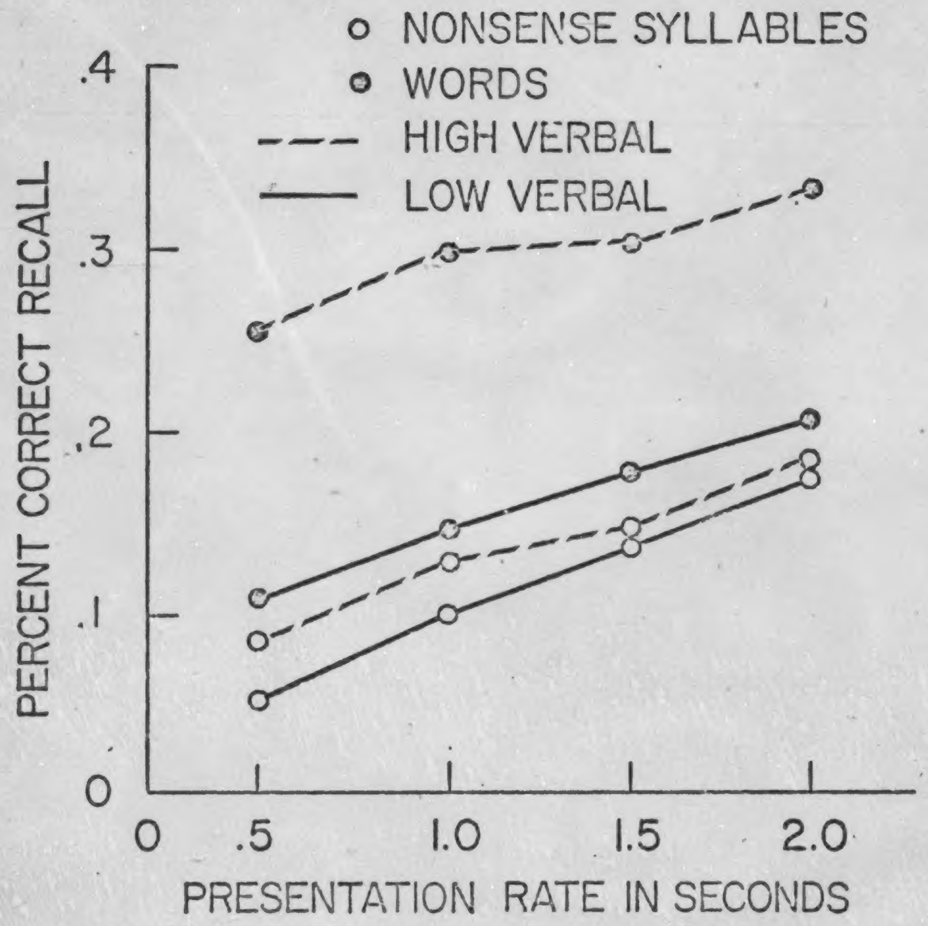
These results suggest that we should make a distinction between coding the stimulus (i.e., associating it with a previously learned pattern) and retaining the code. High verbal ability appears to be associated with one's ability to form the code. The results are even more striking when we examine selected extreme cases. We have examined the strategies for nonsense syllable learning used by persons selected for their extremely high ability to memorize information. Such individuals (mnemonists), who typically have high scores on verbal intelligence tests, use very sophisticated strategies for attaching meaning to nonsense syllables. The chief difference between their strategies and the strategies of unselected students performing the same task (Prytulak, 1971) is that the mnemonists make more use of strategies which associate "nonsense syllables" with meaningful information (Hunt & Love, 1972a; Love, 1973).

Figure 1



Typical results for NI-PI task.

Figure 2



Proportion of word syllables or nonsense syllables recalled as a function of verbal ability and presentation rate.

At the other end of the psychometric scale, mentally retarded (MR) children are notoriously poor memorizers. We have found that MR children have a substantially greater NI-PI difference in the letter identification task than do normal children (Warren & Hunt, 1976). We can sum up these findings by noting that a popular book on "how to improve your memory" advocates strategies which simply cannot work unless one is facile at arousing overlearned codes (Lorayne & Lucas, 1974).

Granted that rapid access to overlearned codes would be useful in memorizing, why should this capability be related to verbal intelligence? Let us consider the nature of the items on a test of verbal comprehension. Many of the items require that the person being tested understand either single words or simple statements. Beginning with the work of Clark and Chase (1972), a considerable number of experiments have been directed at the construction of psychological models of the comprehension process for precisely such statements. This work has been summarized in a theoretical paper by Carpenter and Just (1975). They point out that comprehension requires (a) the establishment of a canonical form of the representations to be compared, regardless of whether these representations derive from linguistic statements or not, and (b) component by component fetching of items from each representation into active memory, their comparison, and the maintenance of a record of agreement-disagreement, logically similar to a parity check comparison in a computer. The complexity of the comparison process can be measured by the number of component checks required. The time for a single fetch can be determined by comparing the time required to analyze a simple statement compared to the time required to analyze a logically identical statement with negative phrasing. To illustrate, suppose that the task (taken from Clark & Chase, 1972) is to determine whether a sentence is true of a picture. In the first condition the sentence PLUS ABOVE STAR is exposed, followed by the picture $\overset{+}{*}$. In the second condition the sentence STAR NOT ABOVE PLUS is exposed, followed by the same picture. In both cases the same response ("true") is required but, according to the Carpenter and Just constituent comparison model, additional memory fetches are required in the second case. If this is so, the difference between the first and second case (Negation verification minus affirmation verification time) should be analagous to the NI-PI difference and should be related to verbal intelligence. Hunt et al. (1975) obtained this result in a contrast between high verbal and low verbal college students. The high-low verbal contrast was extended in an experiment just completed by MacLeod and Hunt. Instead of making the extreme contrast of high vs. low verbals on negative and affirmative sentences, MacLeod and Hunt permitted greater variation of both subject ability and linguistic complexity. They did this by using unselected university students and by presenting four different types of sentences; true affirmative sentences, true negative sentences, false negative sentences, and false affirmative sentences. Carpenter and Just (1975) define a function relating verification reaction time to the four types of sentences. The theoretical justification of the function can be found in their paper. What is important here is that the slope of the function is interpreted as the time to complete a single scan and comparison of two representations in memory. MacLeod and Hunt found that this measure was significantly correlated (-.35) to performance on the Nelson-Denny reading comprehension test. Note that the verification time is measured from the initiation of the picture until the subject's response, and thus the time to read the sentence is not a component of the experimental measure. Also, the correlation is with the slope of a function relating linguistic complexity to reaction time, and not with reaction time itself. Thus this result cannot easily be explained as being due to reading speed. It must be based on the reading comprehension process.

The studies of obligatory coding require that the subject recognize a pattern as being an example of a familiar, named pattern, and then fetch the name into short term (active) memory. What about the structure of active memory itself? Naively, the number of items one can hold in active memory (the *span of apprehension*) ought to be an extremely important parameter of information processing. Miller (1956), after reviewing studies of apprehension for different types of material, but without concerning himself with individual differences, made "7 ± 2" one of the great constants of psychology. Matarazzo (1972, p. 205) observes that the same constancy is true for measures of digit span on the Wechsler Adult Intelligence Scale (WAIS). About 90 percent of the adult population can repeat from five to eight digits correctly. Matarazzo goes on to note, however, that such a range is too small for a psychometric scale. Indeed, digit span has the lowest correlation with full scale IQ of any of the subtests of the WAIS (Robb, Bernardoni, & Johnson, 1972). Matarazzo and others who use the WAIS justify the digit span subtest as an indicator of mental retardation and brain damage, rather than as a guide to intelligence within the normal range. Our own review of the experimental literature (Hunt & Lansman, 1975) also failed to note any startling individual differences associated with the *size* of short term memory. Similarly, our experimental work has failed to produce any striking relation between memory span and other tasks.

Since "size of primary memory" is an attractive concept, we should consider the situation in somewhat more depth. An immediate recall task requires that the subject scan items in memory, organize them into higher order units, and then report the units (Estes, 1972). If this is correct, span of apprehension should be a measure of speed of the manipulation of feature bundles in memory, rather than a measure of the size of a physical short term memory buffer. The apparent short term memory size should depend upon the speed with which codes for higher order units can be recognized, which we have already seen to be related to verbal intelligence, and upon the speed with which active memory can be scanned. Cavanagh (1972) examined the data from some 45 experiments using different stimulus material in the Sternberg (1966) memory scanning procedure, and found a striking linear relation between the reciprocal of span of apprehension and scan rate when measured across different types of stimuli (letters, words, digits, etc.). Apparently the apprehension-scanning relation holds when stimulus content is varied. Does it hold over variation in subjects?

The answer is equivocal. As noted, a good theoretical argument can be made for the position that both short term memory retention and scanning involve chunking information into units and examining the resulting codes. Since these are presumably basic processes in information handling, they should be a component of intelligence. On the other hand, such processes are markedly susceptible to a conscious choice of strategy for memorization, and thus active memory span and, to a lesser extent, scan might show considerable variation both within and across individuals (Estes, 1972; 1974). While such a position makes sense, at present it lacks empirical support. A recent Ph.D. dissertation at the University of Oregon (Lyon, 1975) failed to find evidence for an influence of memory strategies upon memory span in a young adult population. Although one of our early experiments (Hunt *et al.*, 1973) did produce evidence for slow active memory scanning in low verbal college students, compared to their high verbal counterparts, subsequent attempts to replicate this finding have failed. (Indeed, the very high scanning rates observed in that study make us suspicious of the original result.) Sternberg (1975), after an extensive examination of the literature on memory scanning, reported that there are wide individual differences in scan rate within the normal adult population, but that these differences have not been shown to have a consistent relationship either to intelligence or developmental stage. This conclusion is consistent with Matarazzo's observation, based on clinical data, that the absolute difference in memory span within the normal population is not sufficient to make a difference in intellectual functioning, and thus the test of memory span is of marginal usefulness in predicting normal adult performance.

While it is good to have consistent agreement on a conclusion, the conclusion itself is paradoxical. Virtually every modern theory of cognition treats active memory as a bottleneck which limits mental activity. How is it, then, that variations in measures of active memory function fail to show substantial correlations with measures of general intellectual performance? This is a good example of Underwood's (1975) point about individual differences being a crucible for the testing of general cognitive theories. If active memory is such a key feature of human information processing, why aren't individual differences in its functioning among the best predictors of overall cognitive ability?

To add to the complexity of the issue, the whole picture changes if we look at extreme groups! Mental retardates have markedly slow active memory scan rates, especially if the retardation is associated with encephalitic infection and, presumably, resultant brain damage (Harris & Fleer, 1974; Dugass & Kellas, 1974). In a study now in progress we are contrasting the performance of specific etiology mental retardate groups, to determine whether those with presumed brain damage show slower active memory scanning than do equal MA normal children and/or equivalent mental retardates who do not have suspected brain damage. At the other end of the scale, an extremely verbal mnemonist described by Hunt and Love (1972b) had a scan rate of 10 msecs per digit, three to four times as fast as the scan rate usually found for such material in experiments with college students. To gain some idea of the possible relation of scan rate to extremes of span of apprehension, Hunt and Love's mnemonist could memorize a string of 25 digits on a single trial.

The picture presented so far is that short term memory scanning rate, and with it span of apprehension, is an ability which does vary. Within the normal population, however, most people have enough of the ability so that it does not make a marked difference in performance. Superior possession of the faculty is not of great assistance in day to day activity, but may become useful when caught in the psychological laboratory. If this is so, then a relation between scanning rate and intelligence should become apparent within the normal (i.e., university student) population, if the scanning task is (a) difficult enough and (b) requires frequent references to overlearned, long term memory items. One of the experiments reported by Hunt et al. (1975) involved such a task. The problem from the subject's point of view can be grasped by asking yourself "If Monday is 1 and Sunday is 7, how much is Tuesday plus Thursday?" (The answer is "Saturday.") An equivalent, but much harder, problem is "May plus September." College students can learn to solve such problems accurately for, after all, the only knowledge required is the days of the week and the months of the year. By regression analysis of reaction times, we were able to determine scan rates for long term memory lists, and found that in such difficult scan problems the rate of scanning was, indeed, more rapid for high verbal than for low verbal students.

We now examine the relation between long term memory (LTM) and verbal intelligence.

We first consider the storage operation. The simplest paradigm for testing storage in "LTM" (ITM in Hunt's (1971) model) involves information presentation, the presentation of enough interfering material to ensure that retrieval is not from short term memory, and a test trial. In a learning experiment this sequence is repeated until retrieval reaches some criterion for accuracy. The number of trials required to reach criterion is taken as a measure of the rate of storage. An alternative assumption, that information is learned in toto, with probability p, on a given trial would lead one to do the same experiment. In this case the number of trials would also be measured, and used to estimate probability of storage instead of rate. In either case the number of trials required to reach criterion is the *de facto* measurement of storage efficiency.

There appear to be reliable individual differences in storage efficiency. Hunt *et al*. (1973) observed that the number of trials required to learn an arbitrary number-noun list had a correlation of -.51 with verbal ability in college students. This is a fairly crude measure of transfer of information to LTM. More striking results were obtained in a study of "extremely verbal" mnemonists (Hunt & Love, 1972a). These people were selected from respondents to an advertisement offering employment to the "best memorizers in Seattle" and, in order to become participants in the study they had to pass an extremely rigorous test of mnemonic ability. Two of the seven subjects studied were two of the three enrolled University of Washington students with the highest scores on the Washington PreCollege Test's verbal composite. The mnemonists were all extremely rapid learners of paired associate lists. In addition, these subjects participated in the Hebb (1961) repeated digit span task. In this task the subject is ostensibly being assessed for digit span. On each trial *n* digits are presented, where *n* is chosen to be slightly beyond the subject's span of apprehension. The subject then attempts to recall the presented items. Unknown to the subject, every third trial presents the same list of digits. Since the lists intervening between successive presentations of the constant list will effectively fill short term memory (they must, else they would not be beyond the span of apprehension), any improvement on the repeated list must be due to transfer of information into long term memory. Figure 3 shows the performance of mnemonists. For comparison, the figure also shows the results obtained by Hebb (1961) in conditions under which, if anything, the task of learning the constant list should be easier. The mnemonists clearly display superior performance.

Another way of assessing the transfer of information into a storage area where it can be retained at least for minutes is to study the probability of recall of an item that is presented only once, and then tested after a number of other items have been recalled. This is what is done in the "continuous paired associates" paradigm (Atkinson & Shiffrin, 1968). In this task the subject is first presented a pairing of, say, a letter and a number. The next trial presents another pairing. This procedure is continued until all possible stimulus terms have been paired with a response. To illustrate, imagine that the stimulus items are the letters A, B, and C. The initial pairing might be

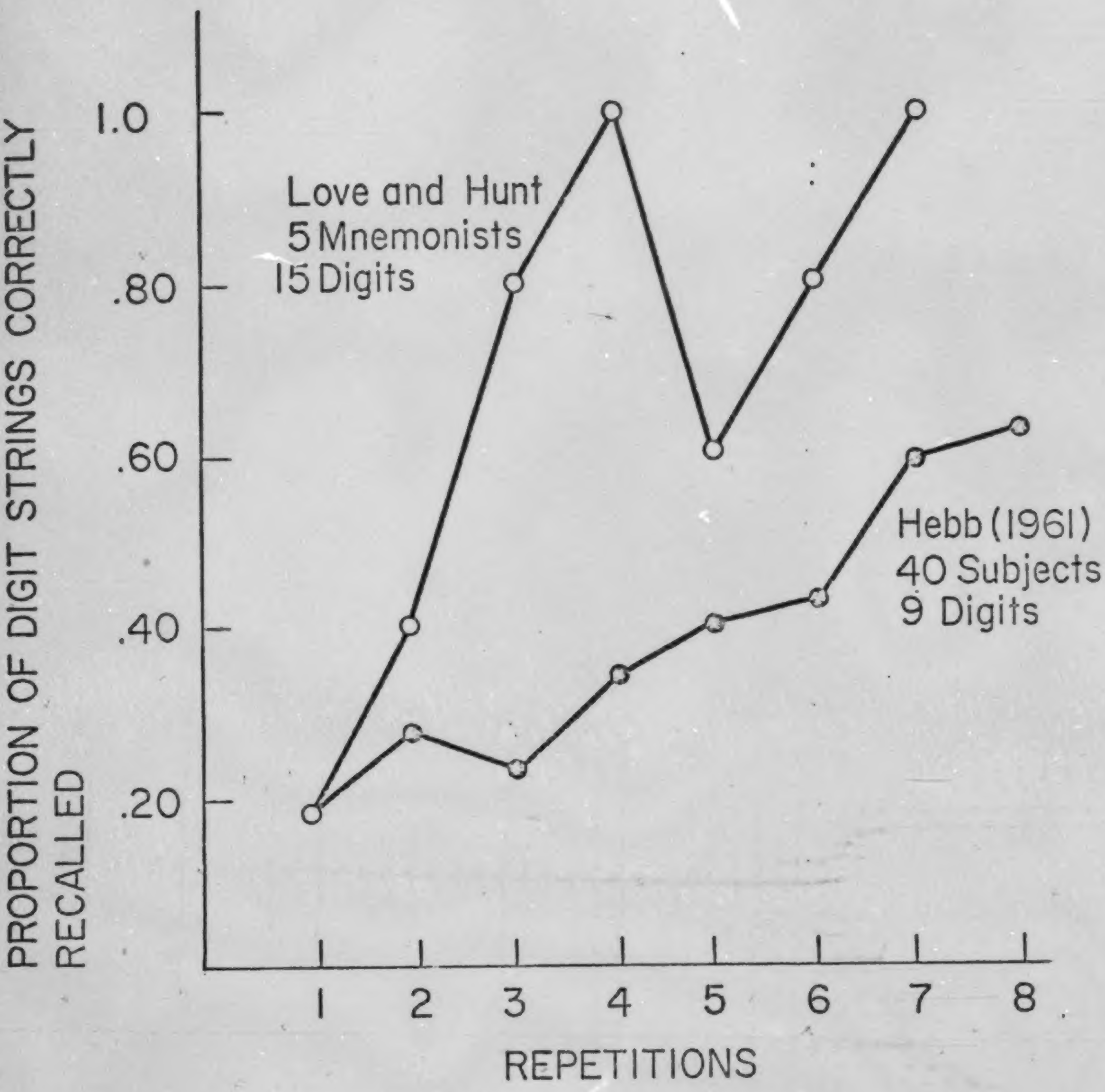
A-3 ; B-6 ; C-14

presented one at a time, in that order. Subsequent trials consist of a *test* phase, in which the stimulus item is presented and the subject must respond with the current response item, and a *study* phase, in which the stimulus is paired with a new response. To continue the example, following the initial pairings we might have

TRIAL 1	test	A - ?	(subject responds)
	study	A - 18	
TRIAL 2	test	C - ?	
	study	C - 2	
TRIAL 3	test	A - ?	(subject responds. 18 is correct.)
	study	A - 45.	

The variable of interest is the probability correct as a function of *lag*, where lag

Figure 3



Comparison of five mnemonists to normals in digit span for repeating digits.

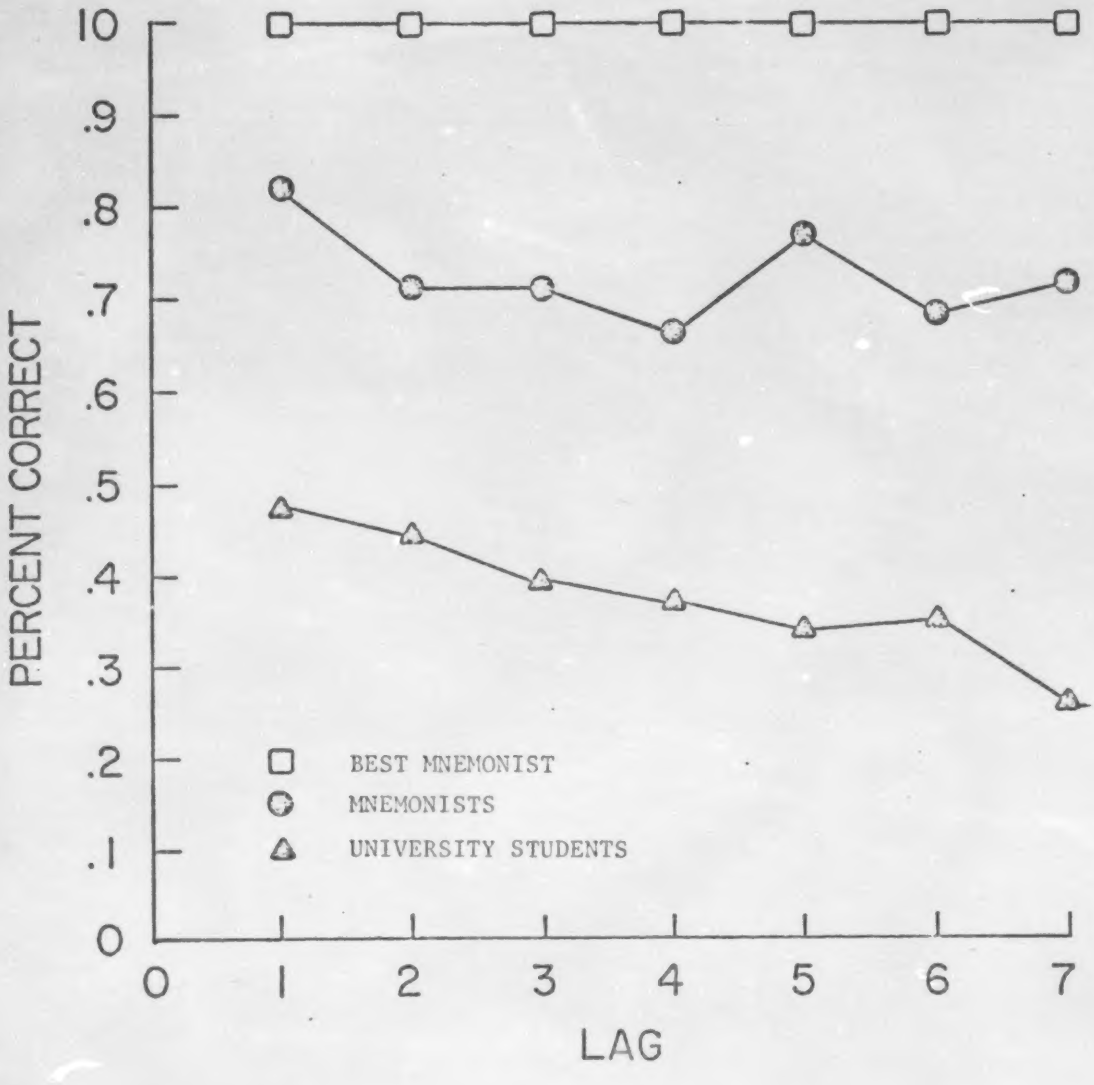
21

is defined as the number of trials intervening between the test of a stimulus and the last study trial using that stimulus. In the above example Trial 3 has a lag of 1. Atkinson and Shiffrin (1968) present an elegant mathematical model of a buffered memory system which well describes their results. For our purposes the important point about this model, and virtually any other in the same situation, is that it assumes that retrieval of information at longer lag values is almost entirely a function of long term memory. (Indeed, Atkinson and Shiffrin's data suggests that this will be the case if lag is much greater than three.) Therefore good performance at longer lags is indicative of either a superior ability to transfer information into long term memory on a single trial, or a superior ability to retrieve information from long term memory, or both. Hunt, Frost, and Lunneborg (1973) found that college students with either high verbal or high quantitative aptitude psychometric scores were superior to low scores on this task, and that the superiority was most evident at higher lag values. The locus of the phenomenon was somewhat unclear. A more striking picture was obtained by Love (1973) in his study of mnemonists. These results are plotted in Figure 4, and demonstrate a clearcut superiority for the mnemonists. Note in particular the "best" mnemonist (not included in the other mnemonist group), who performed without error on an amazing 600 trials, spread over four separate 150 trial sessions!

The problem with these studies is that the locus of the effect is not clear. Are high verbal subjects and mnemonists superior at coding information into long term memory, or at retrieving the information which has been coded? Our further studies indicated that the superiority is in the placing of information into LTM, not in retaining it. On several occasions we have conducted experiments in which subjects of varying verbal ability learned lists of number-noun pairs to a fixed criterion, and were tested for recall up to six weeks later. (Our general method follows the paradigm developed by Nelson (1971), in which savings is used as a measure of retention.) In only one experiment was there even a hint of an interesting effect, and this vanished when the study was repeated. The conclusion is as true for mnemonists as it is for psychometrically defined "high verbals." The same people who have amazing recall for city populations, batting averages, and the contents of stories have quite average recall for meaningless material when original learning is controlled. Interestingly, a similar conclusion has been reached by those working with the mentally retarded. Retardates have notorious memory deficiencies, but these are considerably reduced when an attempt is made to control for original learning (Robinson & Robinson, 1976). These results are consistent with those of others (cf. the review by Hunt & Lansman, 1975) in suggesting that there is relatively little difference in information retention, although there can be considerable difference in information storage.

It is possible that the retrieval process could vary in efficiency, even though individuals did not differ in the information possessed. We have completed one study directed to this question (Hunt, in press, a). College students memorized simple statements containing facts about a group of people, e.g., "Sue is the lawyer," "The lawyer is in the park," and "People in the park can smell the flowers." The next day they returned to the laboratory and answered questions which required retrieval of varying numbers of facts. Continuing the example, the question "Is Sue a lawyer?" requires the retrieval of only one fact, while "Can Sue smell the flowers?" requires the retrieval of three facts. We found that there is a surprisingly strong linear component to the function relating the number of facts to be retrieved to the time required to answer the question. (There is a significant, but much smaller, quadratic component.) Furthermore, the size of this relation varies inversely with verbal intelligence. This finding is of interest because it once again illustrates a relation between verbal intelligence and access to stored information, but it does so in a much more complex situation than the NI-PI or sentence comprehension studies.

Figure 4


PERCENT CORRECT
10
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
0
1
2
3
4
5
6
7
LAG
□ BEST MNEMONIST
○ MNEMONISTS
△ UNIVERSITY STUDENTS

Summarizing briefly, within the population of university students, there appears to be a moderate relation between psychometric verbal intelligence and speed of access to highly overlearned codes, such as the name of a letter. The tasks which demonstrate this are often considered to involve "involuntary" access to memory. Initial results indicate that there is also a relation between psychometric verbal intelligence and speed of access to specific information in long term memory which is accessed on the basis of a conscious effort. There is much less evidence for the latter finding, and further work is clearly in order. Within the normal intelligence range, neither we nor others have found practically any significant associations between intelligence and measures of information manipulation in active memory, apart from the initial act of bringing the relevant codes from permanent memory. There is some evidence that, in cases of either extremely high or low verbal comprehension, active memory parameters (Apprehension, scan rate) are affected. Finally, no one has found significant individual differences in decay of information over long periods of time, provided that original learning is controlled.

Experimental studies II. Verbal comprehension and speech perception skills.

The operational definition of psychometric verbal intelligence does not in any way depend upon manipulation of the spoken language. Most psychologists act upon the assumption that written language is recoded into an internal "acoustic" code and manipulated in that form. If this is so we would expect to find those with high psychometric verbal intelligence scores unusually adept at manipulating speech stimuli. We were encouraged in this belief by Day's (1970) report of a sharply bimodal distribution in the ability to distinguish very small (50 msec) asynchronies in the arrival of dichotically presented speech stimuli. When the arrival times conflicted with what would be expected, given the normal rules of English phoneme order, Day believed that some subjects almost always let their expectations override the actual order of events. While Day did not study intelligence *per se*, she did report that the inability to detect asynchronies was closely related to other measures of performance in tasks requiring manipulation of speech codes. From her reports, one gained the impression that there was a unitary ability to pick apart speech stimuli from their normal context. (To gain some feeling for what we mean, this is what one does in punning.) We wished to explore the nature of this ability further and, in particular, to relate it to verbal intelligence. Our research fell into two lines; verification of Day's findings and the study of the relation between specific speech manipulating abilities and verbal intelligence.

Day's original observations were based on the "fusion" phenomenon. Imagine that two word fragments, i.e., *banket* and *lanket*, are presented dichotically. Some people will report hearing the two stimuli presented, others will "fuse" them into a single word, *blanket*. Fusion can also be identified in a number of other ways, so the illustration is offered as an example only. (See Cutting (1976) for a detailed discussion of different sorts of fusion.) Day reported that her subjects fell into two sharply defined populations of fusers and non-fusers. She found this dichotomy to be reflected in other speech-related skills. In particular, it appeared in a task requiring speech production. She asked subjects to speak in "secret languages," codes based on mutual replacement of phonemes. An example is the /r/ to /l/ secret language, in which *rattle* becomes *lattre* and *Nelson Rockefeller* becomes impossible. Day reported a bimodal distribution in the ability to deal with secret languages, and asserted that it was nearly the same bimodal distribution as that found in the study of fusion.

Day's assertions appear to have been based on an intuitive evaluation of the distribution of responses in a rather small sample. A similar assertion of bimodality was made, with reservations, by Cutting (1975) who combined data from a variety of experimental studies of fusion. We were concerned that Day's samples were too small to evaluate the bimodality assertion, and that Cutting's results on bimodality were obscured by the fact that his assertion was based on a pooled distribution which may have been from populations with different measures of central tendency. We repeated the major aspects of Day's fusion study, using a larger sample (Poltrock & Hunt, 1976) and found that the tendency to fuse stimuli was a reliable measure across individuals, but found no evidence for a bimodal distribution. Cutting (personal communication) provided his raw data, which we reanalyzed by considering the distribution based on differences from the mean fusion within each condition. Professor Cutting has indicated that he believes that this is an appropriate analysis. Again we found no evidence for a bimodal distribution. Day (1976) also made her data available and, in addition, changed her assertion about bimodality. She pointed out that she had run 180 trials, instead of the 60 trials we had run, and that the bimodality in her data appeared only on the last few blocks of trials. (This information was not available in her original reports, nor did she inform us of this prior to being notified of our results.) We have examined Day's data in light of the revised hypothesis. While we agree with her that the data she provided appears bimodal on inspection, there are simply too few subjects for adequate statistical evaluation. Accordingly we now have in progress a larger, and inevitably time consuming experiment, in which a sufficient number of subjects will be run for a sufficient number of trials so that we can (a) settle the bimodality issue and (b) apply some mathematical models to evaluate how the bimodality develops, if indeed it does.

More recently, Day (personal communication) has proposed that the bimodality in speech skills should not be determined solely on the basis of the fusion measure. She feels that a multivariate classification procedure is the best possible. Failing that, she has proposed performance on the secret language task as a good criterion. Testing the multivariate classification proposal would inevitably require a large population, which we hope to do in the experiment in progress. We have made a preliminary analysis of some data obtained using the /r/-/l/ secret language task, and again find no evidence of bimodality, but we are not prepared to make any firmer statement until we have had an opportunity to look into this in more detail.

We do not wish to leave the impression that our work simply questions Day's results. Day reported two findings; that there are reliable individual differences in the skills involved in speech perception and manipulation, and that these skills are bimodally distributed. Our experiments have confirmed the first observation, which, though less publicized, may well be the more important. Further examination of some of these speech skills indicate that they are indeed correlated to verbal intelligence, i.e., to a test of understanding of written speech. We next report our findings on this question.

The ability to distinguish the distinctiveness and temporal order of sounds is basic to speech perception at several levels (cf. Cutting, 1976). Many morphemes are distinguished from each other only by the order of their phonemes. This distinction requires sensitivity to millisecond differences in stimulus onset. Sentence comprehension requires memory for the temporal order of sounds until sufficient information is obtained so that the incoming sentence can be parsed. Even in ordinary speech, the memory period may last for several seconds.

We have conducted two series of experiments related to this question. Hunt et al. (1973) examined the retention of both item and order information in the Peterson and Peterson (1959) paradigm, in which the subject must recall a small number of items after a brief (20-30 sec) retention interval filled with some interfering activity. High verbal subjects exhibited better recall of order information, but there was no difference between high and low verbals in the recall of item information. Note that this is consistent with the finding, previously reported, that high verbal subjects do not differ from low verbal subjects in recall of "nonsense" information without regard to order. The question of retention of information about very slight asynchronies of presentation is more complex. In this situation failure to retain information might be due to failure to detect the original asynchrony of stimulus arrival, or a loss of order information after arrival. It is well known, for instance, that right handed subjects have a bias toward hearing simultaneously presented dichotic speech sounds as having been presented in the order "right ear-left ear" (Cutting, 1976). Hunt et al. (1975) found that the extent of this bias is related to verbal ability... those with higher verbal ability display less of a bias toward hearing "right ear first" and thus are presumably more sensitive to the actual order of presentation.

The relationship between verbal ability and left ear performance for judgment of the order of speech stimuli has been replicated in an experiment using a population representative of the university distribution of verbal intelligence. The same subjects also participated in a temporal order judgment experiment using visual stimuli presented to opposite visual hemifields. Performance on the visual temporal order judgment task was correlated with performance for nonspeech auditory stimuli, but not with performance for speech stimuli. These results suggest that a common mechanism is responsible for judgments of the temporal order of visual and nonspeech stimuli, but a different mechanism limits performance with speech stimuli.

Kristofferson (Kristofferson, 1967; Schmidt & Kristofferson, 1963) suggested that an internal clock limits attentional processes which in turn limit performance in a temporal order judgment task. Others (e.g., Baron, 1969) have suggested that variability in the rate of neural transmission limits discriminations of order. Apparently, psychometric verbal ability is not affected by the limitations of the mechanism, since correlations between psychometric scores and order discrimination performance were near zero for nonspeech stimuli. However, an internal clock could affect performance in tasks which depend on appropriate allocation of effort. One such task requires that subjects read a passage aloud while delayed auditory feedback (DAF) is provided. The DAF task requires that subjects allocate attention to the reading process and avoid processing of the auditory signal. Performance on the DAF task was correlated with order discrimination performance, supporting the hypothesis that an internal clock limits attentional allocation.

A dissertation by Poltrock investigates the nature of the mechanism limiting temporal order judgment performance. In one experiment, subjects were required to judge the order of two visual stimuli and the order of two tones. Performance for the two modalities was highly correlated, $r = .68$, supporting the hypothesis that a common central mechanism is responsible for judgments of order. In other experiments, a few highly practiced subjects judged the order of stimuli presented auditorily, visually, or in both modalities. The results support the existence of a clock which limits performance, but only when stimuli are in different modalities. Judgments of the order of visual stimuli appear to be limited by variability in neural transmission rate, while the order of auditory stimuli were discriminated by a modality-specific mechanism. Apparently, both variability in neural transmission rate and internal clock rate are properties of an individual that limit the individual's sensitivity to order in the environment.

Let us consider alternative ways of testing a person's skill with speech signals, by asking the subject to either read through distortions in the signal or asking him/her to make transformations upon some level of the signal. In delayed auditory feedback (DAF) there is a distortion in the normal temporal order relation cues associated with the production of speech and cues associated with acoustic feedback. The subject has to segregate normally correlated streams of cues on the basis of a temporal order distinction. Several authors have reported that "high verbal" subjects are less affected by DAF than low verbal subjects (e.g., Arens & Popplestone, 1959). We have confirmed this finding in a study which contrasted the top and bottom quartiles of verbal ability, within the university population. The high verbal subjects were able to cope with a DAF interval of 300 msecs, but were disrupted at 150 msecs. The low verbal subjects, by contrast, were disrupted at the longer delay interval. Subsequent work has suggested that this contrast may not be a continuous one, as the correlation between DAF effects and verbal ability in a stratified sample was low (about .20).

We also investigated the relationship between Day's "Secret language" task (Day, 1973) and verbal ability within a stratified sample of university students. The relationship was substantial (.41). Interestingly, in spite of Day's assertions (Day, 1974) that skill with secret languages and ability to cope with DAF are related to the fusion phenomenon, DAF and secret language ability were uncorrelated ($r < .01$).

In summary, we feel that considerable evidence has been obtained for the proposition that skills involved in the perception and production of spoken speech are related to skill in comprehending written speech. It would be of interest to know whether there is a dependency between the development of high orders of skill with the spoken and written language. This question is probably best answered by studies of individual differences in language development and/or disorder. It will be the focus of some of our subsequent research proposals.

Experimental studies III. Psychometric issues.

During the course of our work we have encountered some issues which are of interest to psychometric theory and practice. One is the distinction between verbal comprehension and verbal reasoning. Verbal intelligence tests typically contain subtests which attempt to tap one's knowledge of language facts (e.g., a vocabulary test) and subtests which require reasoning about verbal statements. Some psychometricians (e.g., Guilford, 1967; French, Eckstrom, & Price, 1963) have maintained that these represent distinct abilities. We ran several panel studies in which we obtained data from high school students on (a) psychometrically defined verbal intelligence and (b) performance on a number of information processing tasks. Lunneborg (1974) analyzed this data to determine separate regression equations for predicting verbal comprehension and verbal inference scores from the information processing measures. His results are summarized in Figure 5. As the figure clearly shows, the information processing tests are much more accurate at predicting verbal comprehension subtests than they are verbal reasoning. This is further evidence for the contention that verbal reasoning and verbal comprehension are separate factors. It also provided the impetus for our studies of information processing strategies, rather than mechanics, as the basis for the verbal inference process. We plan to explore this distinction further in our proposed research.

Lunneborg (1975) amplified on these results by calculating regression equations relating psychometric performance to information processing task performance in three separate populations, which can roughly be characterized as "university students, high school students, and bright high school students." He found that the relations between the variables is much weaker in the university/bright high school student population than it is in the "general high school" population. This is not so much due to reduced variance in psychometric test performance as reduced variance in

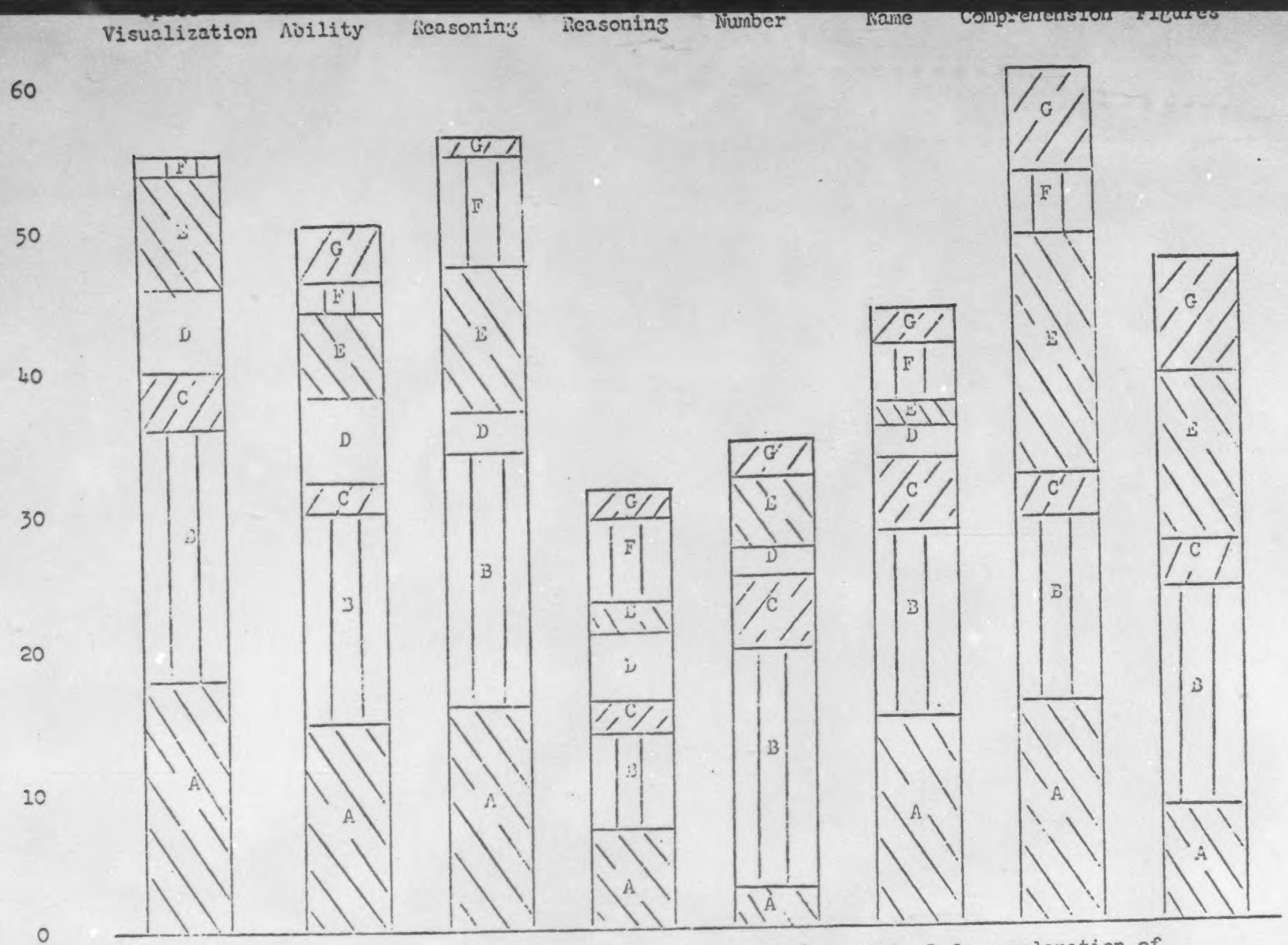


Figure 5. Cumulative Proportion of Variance Accounted For (see Table 5 for explanation of laboratory measures).

Table 3

Predicting Psychometric Performance from Laboratory Measures: Increments to the Squared Multiple Correlation

(Decimal points omitted)

	Verbal Comprehension		Numerical Ability		Space Visualiza		Numerical Reasoning		Verbal Reasoning		Clerical, Number		Clerical, Name		Hidden Figures		
	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	
Motor R T	16	16	15	15	18	18	16	16	07	07	03	03	15	15	08	08	A
Choice Time	22	06	25	10	34	16	28	12	11	04	08	05	20	05	17	09	
Stroop (Wd-*)	25	03	27	02	34	00	31	03	12	01	16	08	27	07	17	00	
Sternberg Intercept	26	01	27	00	36	02	32	01	13	01	19	03	28	01	17	00	
Sternberg Slope	29	03	30	03	36	00	34	02	14	01	20	01	28	00	24	07	B
Pr Corr CRT	32	03	32	02	40	04	34	00	16	02	25	05	33	05	27	03	C
Digit Span, Final	32	00	32	00	42	02	35	01	18	02	26	01	33	00	27	00	
Digit Span, Gain	32	00	38	06	46	04	37	02	21	03	27	01	35	02	27	00	D
Category Score, D L	43	11	43	05	49	03	45	08	23	02	29	02	35	00	37	10	
Ear-Category, D L	49	06	44	01	54	05	47	02	23	00	32	03	37	02	39	02	E
Clustering, B1	53	04	46	02	55	01	55	08	29	06	32	00	41	04	39	00	F
Clustering, B1-R1	55	02	50	04	55	00	55	00	31	02	33	01	42	01	40	01	
B1-B2	59	04	50	00	55	00	55	00	31	00	34	01	43	01	43	03	
B2-R2	60	01	50	00	55	00	56	01	31	00	34	00	43	00	47	04	G

Table 3

Predicting Psychometric Performance from Laboratory Measures: Increments to the Squared Multiple Correlation

(Decimal points omitted)

	Verbal Comprehension		Numerical Ability		Space Visualiza		Numerical Reasoning		Verbal Reasoning		Clerical, Number		Clerical, Name		Hidden Figures		
	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	
Motor R T	16	16	15	15	18	18	16	16	07	07	03	03	15	15	08	08	A
Choice Time	22	06	25	10	34	16	28	12	11	04	08	05	20	05	17	09	
Stroop (Wd-*)	25	03	27	02	34	00	31	03	12	01	16	08	27	07	17	00	
Sternberg Intercept	26	01	27	00	36	02	32	01	13	01	19	03	28	01	17	00	
Sternberg Slope	29	03	30	03	36	00	34	02	14	01	20	01	28	00	24	07	B
Pr Corr CRT	32	03	32	02	40	04	34	00	16	02	25	05	33	05	27	03	C
Digit Span, Final	32	00	32	00	42	02	35	01	18	02	26	01	33	00	27	00	
Digit Span, Gain	32	00	38	06	46	04	37	02	21	03	27	01	35	02	27	00	D
Category Score, D L	43	11	43	05	49	03	45	08	23	02	29	02	35	00	37	10	
Ear-Category, D L	49	06	44	01	54	05	47	02	23	00	32	03	37	02	39	02	E
Clustering, B1	53	04	46	02	55	01	55	08	29	06	32	00	41	04	39	00	F
Clustering, B1-R1	55	02	50	04	55	00	55	00	31	02	33	01	42	01	40	01	
B1-B2	59	04	50	00	55	00	55	00	31	00	34	01	43	01	43	03	
B2-R2	60	01	50	00	55	00	56	01	31	00	34	00	43	00	47	04	G

information processing task performance. In other words, the selection process for students in academic tracks appears to inadvertently select a restricted range of "rapid information processors." This fact should be kept in mind when one considers extrapolating results obtained with a university sample to the general population. Relationships which may be extremely difficult to show in the restricted population may, in fact, be reasonably strong in the general population.

A second study in this series was concerned with the practical application of information processing measures in psychometrics. As we noted earlier, there are powerful cost-effectiveness arguments against using cumbersome experimental procedures as psychometric screening devices..whatever the theoretical justification for the information processing measure. Could we develop a paper and pencil version of some of our information processing measures? We were encouraged by a report by Baddeley (1968) of successful development of a "three minute intelligence test" that is essentially a paper and pencil version of the Clark and Chase (1972) sentence comprehension task. Rose (1974) used the Baddeley test and found moderate (.59) correlations with the verbal intelligence test scores subjects reported having received. We have found a correlation of .28 with verbal reasoning test scores in a university population, which is significant but substantially below the data value by Baddeley. We have also received by an informal report by Price (personal communication) of successful development of paper and pencil versions of the NI-PI task described previously. We now have in progress a study in which paper and pencil versions of the NI-PI task and the sentence comprehension task are being compared to results obtained with the corresponding experimental procedures, and both are being compared to verbal intelligence test scores. So far we have constructed paper and pencil versions of the Chase and Clark sentence validation task, the Posner letter-matching task, and the Shepard and Metzler figure rotation task. We are in the process of validating these paper and pencil measures against the original computer controlled versions. The validation of the Chase and Clark measure is complete, indicating that the computer and the paper and pencil measures are quite highly correlated (RT for affirmatives, .65; RT for negatives, .74; difference, .85) and the two measures show the same pattern of correlations with the Washington PreCollege verbal composits.

Experimental studies IV. Problem solving strategies.

In this sequence of experiments we shift from a microscopic view of cognition as mechanistic information processing to a more macroscopic view of the importance of problem solving strategies. In both computer simulation studies of problem solving and in artificial intelligence research the problem solving task is divided into two stages; the establishment of an internal representation of an external problem and the manipulation of information within that representation. It has been pointed out that the choice of a problem representation will exert a qualitative effect upon problem solving style, even though the same mechanisms may be used to manipulate information once the representation has been chosen (Ernst & Newell, 1967).

This sort of reasoning about the dynamics of problem solving is somewhat foreign to the psychometric approach, where the emphasis is upon a dimensional analysis of the skills required to solve certain problems. Factor analyses of problem solving skills bring with them the implicit assumption that individual performance can be predicted by a linear regression equation based upon a person's values on the "underlying factors of intelligence," and thus have difficulty with discontinuous changes in problem solving behavior. The most extreme form of the implicit assumption is represented by the view that there is a general "intelligence" (g factor) which is loosely characterized as the ability to solve problems. This theory is widespread and, indeed, might well be true. It has important practical consequences. One of these is that if there is a general problem solving skill, then it should be possible to develop a "culture fair" test of this ability. Such a test would be of considerable importance in cross-cultural studies of intellectual functioning. In fact, the Raven Progressive Matrix ("Raven") test (Raven, 1962) has been proposed as such a test (Jensen, 1973; Cattell, 1971).

Proceeding from the computer simulation-computer science view that representation choice implied discontinuous problem solving abilities, Hunt (1973) published a logical analysis of the Raven test. He showed that a "General Problem Solver" algorithm (Newell & Simon, 1972) could be combined with either of two representations to produce rather different patterns of responses to Raven Matrix type items. The analysis was strictly a logical demonstration. No data was offered to show that people actually used different representations. We have now obtained copies of the protocols for the original validation sample for the Raven Matrix test, some 1000 protocols in all. A single linkage method of hierarchical cluster analysis (Johnson, 1967) has been applied to this data, using as a metric the information-theoretic similarity between protocols. (Roughly, this is a technique which specifies closeness between two protocols by noting on which items they agree, and weights agreement on unusual responses more highly than agreement on common responses.) Several of the subsamples have been shown to contain two or three clusters of items, an observation which is more consistent with Hunt's view of the test as subject to different representations, than with the "general factor" explanation of test performance. The latter implies that all protocols should be lumped into a single cluster. While the analyses to date seem to favor some sort of two representation hypothesis, it is not clear that Hunt's hypothesized clusters have been found. Furthermore, the two cluster solution is not found in all samples, nor is it found in a stratified sample of University of Washington students! To further complicate the issue, questions have recently been raised about the statistical reliability of the single linkage technique of cluster analysis (Blashfield, 1976). Further analyses are now underway.

The analysis of the Raven Matrix data was intended to demonstrate that the "multiple representation" view of problem solving was an important one. Now let us look at the other side of the coin, the implication that after a representation has been formed a common problem solving process manipulates that representation. In order to show that this is correct, and an important part of individual intelligence, it is necessary to show that individuals will be consistent in their handling of equivalent representation manipulation problems, even though different representations have been used. To demonstrate this one must construct problems which have different representations, but are equivalent in steps to solution once the representations have been formed. We have constructed two such classes of problems, by amplifying upon an algebraic problem originally due to Malin (1973). In Malin's version of the problem the subject is presented with a set of algebraic identities, such as

Distance = speed X time
Cost = fuel x price
Fuel = distance / m.p.g.

A problem is defined by designating certain variables as known, and asking the subject to determine whether a target variable can be derived. (The subject does not need actually to derive the variable, merely to determine that it can be derived.) Using the example, if DISTANCE and COST are known, can FUEL be derived? The answer is "no," because PRICE and M.P.G. are not given.

The formal process of deriving an answer can be depicted as a search through a problem reduction graph (Nilsson, 1971). Continuing the example, FIND FUEL will generate the problems (FIND COST, PRICE) and (FIND DISTANCE, M.P.G.). Each subproblem will be either necessary or sufficient for the solution of the problem which generated it. Various algorithms can be applied to ensure an efficient search through the resulting problem reduction graph.

It is possible to generate figural isomorphs of Malin's algebraic problems. The isomorphs are somewhat akin to jigsaw puzzles. An example is shown in Figure 6, which demonstrates how jigsaw pieces can be fitted together to represent an equation. In the jigsaw isomorph of a problem the subject is given a certain number of pieces, corresponding to the "givens" of the algebraic problem, and asked if an assemblage can be created which will include a target piece. Note that the problem isomorphism is complete. For every equation there is a corresponding assemblage, and for every derivation there is a corresponding fitting of a new piece into a previously constructed assemblage, and vice versa. This is true in spite of the strikingly different perceptual aspect of the problem. (We have never had a subject recognize that the problems are the same.) Nevertheless, over 26 subjects, we obtained a correlation of .76 for "average number of moves" in several problems of each type. This is much higher than the correlations usually reported between figural and symbolic problem solving performance, but the previous studies have varied both the representation and the search for solution after the representation has been formed.

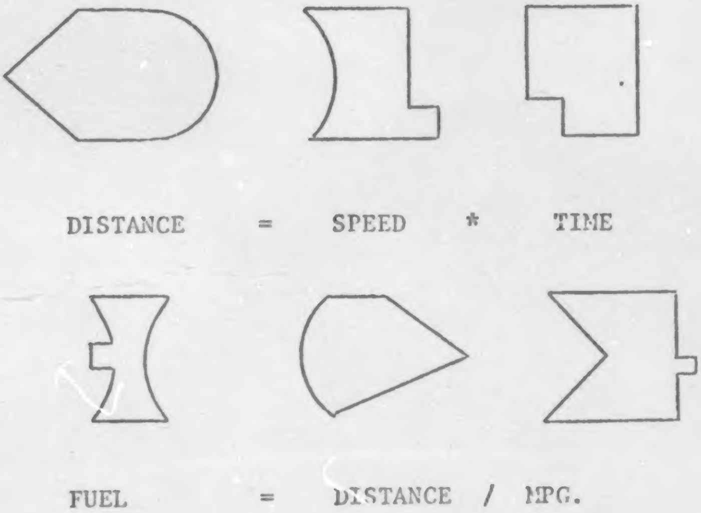
Experimental studies V. Application of information processing measures to applied problems in cognitive measurement.

Although this is a basic research project, we are understandably interested in possible applications of our work to applied problems. Two such efforts were suggested to us. One deals with the establishment of fine grained profiles of mental deficiency associated with certain types of retardation. The other represents an interesting use of mental measurements to establish intra-individual fluctuations in cognitive efficiency. These applications were originally suggested by the University of Washington Child Development and Mental Retardation Center and by the National Institute of Nervous Disease and Stroke. Both these institutions have provided the major support for the projects to be described, although some funds from this grant have been used, mostly to provide equipment support and some time of project personnel. As is often the case, these forays into applied research have suggested additional theoretical questions, which will be pursued in our proposed research section.

We first describe our studies of specific etiologies of mental retardation.

"Mental retardates" are not homogeneous groups of individuals. The state of mental retardation is a constellation of behaviors displayed by persons whose case histories may vary from a record of social deprivation, a history of mental retardation in the family, specific environmental insult, such as anozia during birth, or specific physiological abnormalities, such as the genetic abnormalities associated with Down's, Turner's, or Kleinfelter's syndromes. In spite of such varied etiology, mental retardation is often treated as a unitary phenomenon which can be adequately summarized by a score on a global intelligence test. The reader who doubts this is invited to examine the descriptions of "level of retardation" provided in the typical experimental report. We wager that nine of ten such reports provide only IQ or MA information. Our theoretical approach implies that this is inadequate. We would expect particular types of structural or physiological damage to be reflected

Figure 6


DISTANCE = SPEED * TIME
FUEL = DISTANCE / MPG.

An example of jigsaw and algebraic problem isomorphs.

by specific patterns of behavior. To put the matter somewhat differently, many studies of mental retardation are based on the (implicit or explicit) assumption that the hallmark of mental retardation is a failure to develop mental skills. It is for this reason that investigators often look at behavior in tasks which, in the normal population, show a developmental trend. We agree that this approach may well be adequate for some types of retardation, especially those associated with "cultural-familial" retardation, but that the picture might be quite different in the case of specific etiologies. Indeed, there is evidence that this is true in some areas of cognition. X chromosome defects appear to lead to severe, selective impairment of spatial reasoning (Money, 1968). We ask whether similar selective defects of memory may be associated with other specific etiologies.

Warren and Hunt (1976) contrasted the performance of an undifferentiated retardate group to an equivalent MA group of Prader-Willi syndrome cases. The Prader-Willi syndrome is described by mild mental retardation, arrested sexual development, and, most strikingly, extreme obesity and an obsessive desire for high caloric food. (Prader-Willi cases can maintain weight on 900 calories per day.) Although the etiology of the disease is not definitely established, the behavioral and physiological observations are strikingly parallel to the observations of animals with experimentally produced damage to the hypothalamus. Because of the evident general involvement of the limbic system in learning and memory, we might expect that the Prader-Willi group to be unusually deficient in this area.

Warren and Hunt compared Prader-Willi children, equal MA normal children, and an undifferentiated retardate, equal MA group in their performance on a continuous picture recognition task. This task was chosen because picture recognition minimizes the development of rehearsal strategies. Adolescent retardates give near-normal performance on such a task, which has lead Brown (1972) and others to suggest that the defect in retardate memory is largely due to an inability to generate rehearsal strategies. (Brown, as is common, did not differentiate between different etiologies of retardation.) Warren and Hunt replicated Brown's results with undifferentiated older retarded children, but found Prader-Willi retardates to be impaired in picture recognition. An even more interesting finding emerged from consideration of the "developmental trend" in this task. Normal children are very good at picture recognition almost as soon as the task can be explained to them. Warren and Hunt found a correlation of .86 between CA and picture recognition performance in undifferentiated retardates, which is clearly what would be expected from an "arrested development" view of retardation. There was virtually no correlation between performance and age in the Prader-Willi group (r = -.14). At the least, this seems to warrant caution in asserting an "arrested development" view of cognitive functioning in all retardates.

In another part of their study, Warren and Hunt contrasted the performance of the two retardate groups and an equal CA control group on the NI-PI task, described previously, which served as a measure of access to overlearned material. The main motivation for doing this was to show that the Prader-Willi group was not uniformly more retarded than the undifferentiated group on all measures of cognitive functioning, a fact which must be established to maintain that the degree of retardation was the same between groups. The two retardate groups were equivalent, but both had a memory access time considerably longer than that of the (probably slightly superior intelligence) "normal" children. This is interesting because we can look at the retardate-normal contrast as being the opposite pole from the normal-superior contrasts made in the studies using university students. It is encouraging to find consistent results with these quite different populations. Firm conclusions, however, will depend upon obtaining much more data on the normal developmental course of the various information processing measures.

The next step in this research is clearly to determine what, if any, orderly relation can be found between specific defects in information processing and specific etiologies of retardation. We point out that this question is of interest in the development of a theory of cognition in general, as it provides a way (on occasion) to tie human cognition to biologically defined differences between individuals. The work is also of some interest in mental retardation, even though it is outside the "arrested development" viewpoint now dominant in that field.

Our second series of application studies differs from our usual design in that we became interested in intra-individual fluctuation of cognitive efficiency, as opposed to inter-individual differences. The impetus for the studies came from the observations of Dr. Anatol Dekaban, of the NINCDS staff. Dr. Dekaban pointed out that patients suffering from severe epilepsy, who are consequently heavily medicated, often report that the medication leaves them physically and mentally exhausted. In patient management it would be desireable to calibrate such a side effect. Conventional intelligence tests do not provide an adequate measurement for several reasons, not least of which is that they cannot be given repeatedly over short intervals. Would it be possible to use information processing measures of cognitive functioning to assess the undesirable effects of the medication? If so, this would both be useful in clinical practice and provide an interesting tool for the study of other medical experimental manipulations which might have effects on cognition. Dr. Dakaban joined us for a year, to conduct joint studies of this question. Severe epileptic patients are being tested on a battery of measures of cognitive effectiveness, derived from our theoretical model. At the same time their level of medication is being manipulated systematically. In two parallel experiments, non-epileptic volunteers are going through the same regimen, some medicated and some not medicated, permitting us to evaluate the effects of the drug per se and in interaction with the pathological process. As a very large amount of data has been gathered, and the study is not yet completed, we can provide only partial findings.

The basic plan of the epilepsy study was a within subjects design, varying level of phenobarbital medication. All our subjects were extreme epileptic cases, and thus had an existing level of medication. They were tested on eleven different cognitive tasks. Week 1 testing was carried out under the medicinal level, week 2 testing was a replication of week 1, but under an increased level of phenobarbital. Ideally, level of drug dose would be counterbalanced. This was not possible, however, because of the expense involved. Phenobarbital has a 7 day half-life, and a counterbalanced design would have required that "high dose first" subjects be retained in a hospital for an additional week or more. Therefore two control groups were run. A "normal-drug" group consisted of non-epileptic volunteers who performed the tasks for a week, without a drug, and in the subsequent week performed while under the influence of phenobarbital. This group permitted assessment of the effect of the drug without neurological disorder. Finally, a no-drug control group completed two weeks work on the task without medication. This provided a control for improvement due to learning.

The task battery itself contained tasks intended to tap the various aspects of active and long term memory which have already been mentioned. Data analysis for this project is a substantial problem, due to the volume of data collected. We can note, at this time, some promising trends. The normal controls showed steady improvement over two weeks. Epileptic patients and drugged controls did not. Both drug groups showed a specific deficit in performance under "higher drug" conditions on those tasks which involved access to memory. On the other hand, active memory tasks, such as the Sternberg memory scanning paradigm, seem not to have reacted to the drug. We stress again that the data analyses have not been completed. What we have seen, so far, appears to complement the research on differences in information processing across individuals.

The present state of affairs

The past five years of research have clearly been successful in a gross sense. We have been able to demonstrate that a connection between the information processing and individual differences approaches to cognition is possible. This is no mean accomplishment. Historically, the attempt to connect a learning-oriented experimental psychology to psychometrics did not succeeed, nor has a perception-based approach been any more successful. Our efforts have not gone unnoticed. Invited addresses have been given before the American Educational Research Association and the American Psychological Association. Recent Ph.D. dissertations by Lyon (1975) and Sternberg (in press), done elsewhere, reflect a new interest in individual differences as more than a source of error in otherwise well run experiments. There has been a provocative attempt to analyze psychometric tests in terms of the cognitive functions involved (Carroll, 1976). We do not for a moment claim all the credit for this reapproachment between experimentalists and psychometricians, but we believe we have assisted it.

At a more prosaic level we have established a fact. Psychometrically defined verbal intelligence is associated with rapidity of access to well learned information. The evidence for this assertion is strong when access is defined as access to "obligatorally coded" material, such as the name codes for letters or words. The evidence is present but weaker for voluntary retrieval of episodic information. We have also obtained evidence that the high score on a written verbal intelligence test display superior skills in certain aspects which relate to spoken speech perception. Roughly, these skills appear to be associated with the ability to override biases toward reacting to what is expected in speech, and instead detecting what really happened.

These findings are important in themselves because they tell us something about the dynamics of individual differences in the comprehension of verbal material. At a more general level, we have established the existence of individual differences in the mechanisms of information processing which are associated with intelligence, and which cannot reasonably be said to depend upon the possession of knowledge. The detection of differences in information handling mechanisms is essential if we are to go beyond the level of descriptive correlational analyses in explaining individual difference in cognition.

There are two things which we have not yet done. Most important, we have not proposed a cognitive theory from which individual differences can be derived. There is little point in "explaining the intelligence test score" as an exercise in itself. Correlational studies such as ours are a useful first step, for they indicate the variables that must be involved in a process theory of cognitive differences. To take the next step we must relate the experimental variables to each other. In addition, we must incorporate the more recent findings in general experimental psychology into our model. These steps are taken, albeit only partially, in the following section on Proposed Research.

is quite small.

Facilities Available

Our laboratory includes eight computer-controlled testing stations. Each of the stations is isolated and sound attenuated; six are in the large room housing the computer and two are in ancillary rooms, which can be observed through one-way mirrors. Each station is equipped with a Tektronix scope for visual displays, and headphones for auditory presentation of material. Subjects' responses are made through keyboard consoles at each station; voice keys may be easily connected when desired.

A Nova 820 computer with 24K memory controls stimulus presentations and monitors subjects' responses. Peripheral hardware includes fixed and removable disks, Linc tapes, a high-speed printer and an Imlac mini-computer (interfaced to the Nova) which can be used for text and graphics displays. In addition, our computer is connected to the main departmental computer, allowing use of the mag tape system and high-speed paper-tape readers. Large data analyses and simulations may be done on the campus computer, a CDC-6400; we are currently arranging an interface and use mag tapes for inter-computer communication in the meantime.

Our laboratory also includes many other useful pieces of equipment. We have voice keys, tape recorders (both two- and four-track), slide projectors, signal generators, cycle timers, and other experimental apparatus available, in addition to other departmental facilities. Two additional rooms are provided for data analysis and for special experiments.

A unique software system has been developed to provide real-time control of experiments in our laboratory by Burkhardt and his associates (Burkhardt, 1976a,b; Burkhardt, Palmer, & Waddington, 1976). The system, now fully operational, features a high-level, user-oriented programming language called PLE-1. The system is well-documented and provides the background necessary for all input, output and timing operations. Using the PLE language, we have been able to develop, in a matter of hours, programs that control experiments similar to those described in this proposal.

Arrangements for participants in the experiments will be quite straightforward. The Psychology Department maintains a full time liaison officer to assist in such arrangement. University students will be solicited through the mail, the selections being made from testing records in the Bureau of Testing (Dr. Clifford E. Lunneborg is the Director of the Bureau). We have previously had (and are presently renewing) connections with the Seattle School Board, so as to obtain younger subjects when that is desireable.

References

Abelson, R. The structure of belief systems. In Schank, R. and Colby, K. (ed.) Computer models of thought and language: San Francisco, Freeman, 1973

Amarel, S. On machine representations of problems of reasoning about actions. The missionaries and cannibals problem. in Michie, D. (ed.) Machine Intelligence 3 Edinburgh. Edinburgh U. Press, 1968

Anderson, J. and Bower, G. Human Associative Memory. New York: Academic Press, 1973

Arens, J. and Popplestone, J.A. Verbal facility and delayed speech feedback. Perceptual and Motor Skills 1959, 9, 270.

Atkinson, R. C. and Shiffrin, R. M. Human memory: A proposed system and itsccontrol processes. in Spence, K. and Spence, J. (ed.) The psychology of learning and motivation (vol. 3). New York: Academic Press, 1968

Attneave, F. and Arnoult, M. A quantitative study of shape and pattern perception. Psychol. Bull. 1956, 53, 452-471

Baddeley, A.D. A 3 minute reasoning test based on grammatical transformation Psychon. Science 1968, 10,341-342

Barclay, J. R. The role of comprehension in remembering sentences. Cogn. Psychol. 1973, 4, 229-254

Black, J. and Carter, R. Association between aortic stenosis and facies of severe infantile hypercalcaemia Lancet, 1963, 2, 745-749

Blashfield, R.K. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchial methods. Psychol. Bull. 1976, 83, 377-388

Bloom, B. and Broder, L. The problem solving processes of college students. Chicago: U. Chicago Press, 1950

Bower, G. Cognitive Psychology: An introduction. In Estes, W. K. (Ed.) Handbook of learning and cognitive processes. Hillsdale, N.J. Erlbaum Assoc. 1975

Bransford, J.D. and Franks, J.J. The abstraction of linguistic ideas. Cogn. Psychol. 1971, 2, 331-350

Brown, A. Context and recency cues in the recognition memory of retarded children and adolescents. *Am.J. Mental Deficiency* 1972, *77*, 54-58

Carpenter, P.A. and Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychol. Rev.* 1975, 82, 45-73

Cattell, R. B. *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin, 1971

Cavanagh, J.P. Relation between the immediate memory span and the memory search rate. *Psychol. Rev.* 1972, *79*, 525-530

Clark, H. and Chase, W. On the process of comparing sentences against pictures. *Cogn. Psychol*. 1972, *3*, 472-517

Collins, A. and Loftus, E. A spreading activation theory of semantic processing. *Psychol. Rev.* 1975, *82*, 407-428

Cooper, L. and Shepard, R. Chronometric studies of the rotation of mental images. in Chase, E. (ed.) *Visual information processing* New York: Academic Press, 1973

Cote, B., Nix, S., and Frost, N. Paper presented at Western Psychol. Association, 1971

Court, J. H. *Researchers' Bibliography on Ravens Progressive Matrices and Mill Hill Vocabulary Scales* (3rd ed.) Flinders U. of S. Australia 1974

Cutting, J. E. Aspects of phonological fusion. *Journal of Expt'l Psychol.: Human perception and performance*, 1975, *1*, 105-120

Cutting, J. E. Auditory and linguistic processes in speech perception: inferences from six fusions in dichotic listening. *Psychol. Rev.* 1976, *83*, 114-140

Cutting, J.E. Personal communication, April 1976

Day, R.S. Temporal order judgements in speech: Are individuals language bound or stimulus bound. Haskins Laboratories status report 1970, SR-21/22 71-87

Day, R.S. On learning "secret languages" Haskins Laboratories status report 1973, SR 34

Day, R.S. Address given at American Psychological Symposium on Individual Differences in Cognition. (Div. 3) American Psychological Association meeting, New Orleans, 1974

Day, R.S. Distribution of fusion scores as a function of number of trials. Unpublished manuscript, April 1976

Dugas, J.L. and Kellas, G. Encoding and retrieval processes in normal children and retarded adolescents. *J. expt'l Child Psychol*. 1974, *17*, 177-185

Dunn, H. The Prader-Labhart-Willi syndrome. Review of the literature and report of nine cases. *Acta Pediat. Scan.* 1968, *186*, (supplement) 1-38

Ernst, G. and Newell, A. Some issues of representation in a general problem solver. *Proc. Spring Joint Computer Conference* AFIPS 19, 1967

Ernst, G. and Newell, A. *GPS: A case study in generality and problem solving*: New York: Academic Press, 1968

Estes, W.K. The cognitive side of probability learning. *Psychol. Rev.* 1976, *83*, 37-64

Estes, W.K. An associative base for coding and organization in memory. in Melton,A. and Martin, E. (ed.) *Coding processes in human memory* Washington: Winston-Wiley, 1972

Estes, W.K. Learning theory and intelligence. *Amer. Psychol*. 1974, *29*, 740-749

Fanconi, G., Giradet, P. Schlesinger, B., Butler, N. and Black, J. Chronische Hypercalcamie, Kombinert mit Osteosklerose, Hyperazotomie, Minderwuchs, and Kongenitalen Missbildungen. *Helve. Paediat. Acta*. 1952, *7*, 314-349

Forbes, A. Item analysis of the advanced matrices. *British J. Psychol*. 1964, *34*, 223=236

Forbes, G., Bryson, M. and Manning, J. Impaired calcium homeostasis in the infantile hypercalcemic syndrome. *Acta Pediatr. Scand*.1972, *61*, 305-309

French, J., Ekstrom, R.B. and Price, L. *Manual for kit of reference tests for cognitive factors*. Princeton, N.J. Educational Testing Service, 1963

Friedman, A. and Bourne, L.E. Jr. Encoding the levels of information in pictures and words. *J. exp. Psychol.: General* 1976, *105* 169-190

Gibson, E.J., Gibson, J.J., Pick, A.S., & Osser, H. A developmental study of the discrimination of letter-like forms. *Journal of Comparitive Physiological Psychology*, 1962, *55*, 897-906.

Gopher, D. & Kahneman, D. Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills*, 1971, *33*, 1335-1342.

Groen, G.J. & Parkman, J.M. A chronometric analysis of simple addition problems. *Psychological Review*, 1972, *79*, 329-343.

Groen, G.J. & Poll, M. Subtraction and the solution of open sentences problems. Report 72-70. Psychology Department, Carnegie-Mellon University, Pittsburgh, PA, 1972.

Guilford, J.P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.

Guilford, J.P. & Hoepfner, R. *The analysis of intelligence*. New York: McGraw-Hill, 1971.

Hall, B., & Smith, D. Prader-Willi syndrome. *J. Pediat.*, 1972, *81*, 286-293.

Harris, G.J. & Fleer, R.E. High speed memory scanning in mental retardates: Evidence for a central processing deficit. *Journal of Experimental Child Psychology*, 1974, *17*, 452-459.

Hartigan, J.A. *Clustering algorithms*. New York: Wiley, 1975.

Hays, W.L. *Statistics for the social sciences*. New York: Holt, Rinehart, and Winston, 1973.

Hebb, D.O. Distinctive features of learning in the higher animal. In J. Delafresnay (Ed.), *Brain mechanisms and learning*. London: Oxford University Press, 1961.

Horn, J. Organization of abilities and the development of intelligence. *Psychological Review*, 1968, *75*, 242-259.

Horn, J.L., & Catell, R.B. Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 1967, *26*, 107-129.

Hunt, E. What kind of computer is man? *Cognitive Psychology*, 1971, *2*, 57-98.

Hunt, E. Quote the Raven? Nevermore! In L. Gregg (Ed.), *Knowledge-Cognition*. L. Erlbaum Assoc., 1974.

Hunt, E. Varieties of cognitive power. In L. Resnick (Ed.), *Nature of Intelligence*. Hillsdale, N.J.: Erlbaum Assoc., 1976.

Hunt, E. We know who knows, but why? In R. Anderson & R. Spiro (Eds.), *Learning and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum Assoc., in press (a).

Hunt, E. Imageful thought. Proceedings of the Conference on Semantic Factors in cognition, in press (b).

Hunt, E.B. *Artificial Intelligence*. New York: Academic Press, 1975.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In C. Bower (Ed.), *Advances in learning and motivation*, *7*. New York: Academic Press, 1973.

Hunt, E. & Lansman, M. Cognitive theory applied to individual differences. In W. Estes (Ed.), *Handbook of learning and cognitive processes*, *1*. Hillsdale, N.J.: Erlbaum Assoc., 1975.

Hunt, E. & Love. L.T. How good can memory be? In A. Melton & E. Martin (Eds.), *Coding processes in human memory*. V.H. Winston & Sons, 1972 (a).

Hunt, E. & Love. T. The second mnemonist. Address given at APA conference, 1972.

Hunt, E., Lunneborg, C., & Lewis, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, *7*, 194-227.

Hunt, E. & Poltrock, S. The mechanics of thought. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, New Jersey: L. Erlbaum Assoc., 1974.

Jensen, A.R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1-123.

Jensen, A.R. Hierarchical theories of mental ability. In W.B. Dockrell (Ed.), *On intelligence*. London Methoen, 1970.

Jensen, A. R. *Educability & group differences*. New York: Harper & Row, 1973.

Johnson, D.M. *Systematic introduction to the psychology of thinking*. New York: Harper & Row, 1972.

Johnson, S.C. Hierarchical clustering schemes. *Psychometrika*, 1967, *32*, 241-254.

Jones, K., & Smith, D. The Williams elfin facies syndrome: A new perspective, in press.

Kahneman, D. *Attention and performance*. New York: Academic Press, 1973.

Kahneman, D., Ben-Ishai, R., & Lotan, M. Relation of a test of attention to road accidents. *Journal of Applied Psychology*, 1973, *58*, 113-115.

Kantowitz, B.H. Double stimulation. In B.H. Kantowitz (Ed.), *Human information processing*, Hillsdale, N.J.: Erlbaum Assoc., 1974.

Kite, E.S. Translation of A. Binet and T. Simon's, *The development of intelligence* in children. Baltimore: Williams and Wilkins, 1966.

Klatzky, R.L. & Stoy, A.M. Using visual codes for the comparison of pictures. Memory and Cognition, 1974, 2, 727-736.

Kristofferson, A.B. Successiveness discrimination as a two-state quantal process. Science, 1967, 158, 13337-1339.

Lanvin, D.E. The prediction of academic performance. New York: Russell Sage Foundation, 1965.

Loftus, E.F. & Cole, W. Retrieving attribute and name information from semantic memory. Journal of Experimental Psychology, 1974, 102, 1116-1122.

Lorrayne, H. & Lucas, J. The memory book. New York: Ballantine Books, 1974.

Love, L.T. Location and analysis of exceptional memory. Unpublished master's thesis. Psychology Department, University of Washington, 1973.

Lunneborg, C.E. Address given at American Psychological Association (Div. 3). Symposium on individual differences in cognition. New Orleans, 1974.

Lunneborg, C. Choice reaction times and psychometric performance revisited. Technical Report 76-7. Educational Assessment Center, University of Washington, 1975.

Lyon, D.R. Source of individual differences in digit span. Unpublished doctoral dissertation, University of Oregon, 1975.

Malin, J.E.T. An analysis of strategies for solving certain substitution problems. University of Michigan, Human Performance Center, Technical Report No. 40, April, 1973.

Matarazzo, J.D. Wechsler's measurement and appraisal of adult intelligence. (5th ed.) Baltimore: Williams and Wilkins, 1972.

Miller, G. The magic number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Moeser, S.D. Inferential reasoning in episodic memory. Journal of Verbal Learning and Verbal Behavior, 1975, 15, 193-212.

Money, J. Cognitive deficits in Turner's syndrome. In S. Vandenberg (Ed.), Progress in human behavior genetics. Baltimore: Johns Hopkins Press, 1968.

Nelson, T.O. Savings and forgetting from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 568-576.

Newell, A. & Simon, H. Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1972.

Nillson, N. Problem solving methods in artificial intelligence. New York:

McGraw-Hill, 1971.

Norman, D. & Bobrow, D. On data limited and resource limited processes. *Cognitive Psychology*, 1975, *7*, 44-64.

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.

Paris, S.G. & Lindauer, B.K. The role of inferences in children's comprehension and memory for sentences. *Cognitive Psychology*, 1976, *8*, 217-227.

Paris, S.G., & Mahoney, G.J. Cognitive integration in children's memory for sentences and pictures. *Child Development*, 1974, *45*, 633-642.

Peterson, L. & Peterson, M. Short term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, *58*, 193-198.

Pipes, P., & Holm, V. Weight control of children with Prader-Willi syndrome. *Journal of American Diet*. Association, 1973, *62*, 520-524.

Poltrock, S.E., & Hunt, E. Individual differences in phonological fusion and separation errors. *Journal of Experimental Psychology: Human Perception and Performance*, (in press).

Posner, M.I., Boies, S.J., Eichelman, W.H., & Taylor, R.L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, *79*, #1, pt. 2, 1-16.

Posner, M.I. & Mitchell, R.E. Chronometric analysis of classification. *Psychological Review*, 1967, *74*, 392-409.

Posner, M.I. & Snyder, C. Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition*. Erlbaum Assoc., 1975.

Prader, A., Labhart, A., & Willi, H. Ein syndrom von adipositas kleinwuchs, krytorchismus, und oligophrenia nach myatonieartigem zustand im neugeborenenalter. *Schweiz. Med. Wochenschr.*, 1956, *86*, 1260.

Price, J. Person Communication, 1975.

Prytulak, L. Natural language mediation. *Cognitive Psychology*, 1971, *2*, 1-56.

Pylyshyn, Z.W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1-24.

Raven, J.C. *Advanced progressive matrices*. London: H.K. Lewis, 1962.

Rips, L.J., Shoben, E.J., & Smith, E.E. Semantic distance and the verification of semantic relations. *Journal of verbal learning and verbal behavior*, 1973, *12*, 1-20.

Robb, G.P., Bernardoni, L.C., & Johnson, R.W. *Assessment of individual mental* *ability*. Scranton, PA: Intext Educational Publishers, 1972.

Robinson, N. & Robinson, H. *The mentally retarded child* (2nd ed.). New York McGraw-Hill, 1976.

Rose, A.M. Human information processing: An assessment and research battery. AD 785 411, National Technical Information Service, 1974.

Schank, R. *Conceptual information processing*. New York: American Elsevier, 1975.

Schmidt, M.W., & Dristofferson, A.B. Discrimination of successiveness: A test of a model of attention. *Science*, 1963, *139*, 112-113.

Schwartz, S. Individual differences in information processing: Verbal ability and memory encoding processes. Invited paper presented at the annual meeting of the Midwestern Psychological Association. Chicago, May, 1976.

Shepard, R. & Metzler, J. Mental rotation of three dimensional objects. *Science*, 1971, *171*, 701-703.

Simon, H.A. Identifying basic abilities underlying intelligent performance of complex tasks. In L.B. Resnick (Ed.), *The nature of intelligence*, New York: Wiley, 1976.

Simon, H. & Hayes, J. The understanding process: Problem isomorphs. *Cognitive Psychology*, 1976, *8*, 165-190.

Smith, E., Shuben, E., & Rips, L. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214-241.

Spearman, C. *Abilities of man*. London: Methven, 1927.

Sperling, G. Personal communication. October, 1976.

Sperling, G., Budiansky, J., Spivak, J.G., & Johnson, M.C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science*, 1971, *174*, 307-311.

Sternberg, R. *The componential analysis of human abilities: Intelligence, information processing, and analogical reasoning*. Hillsdale, N.J.: Erlbaum Assoc., in press.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, *153*, 652-654.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, *57*, 421-457.

Sternberg, S. Memory scanning: New findings and correct controversies. *Quarterly Journal of Experimental Psychology*, 1975, *27*, 1-32.

Thurstone, L.L. *Primary mental abilities*. Chicago: University of Chicago Press, 193

Townsend, J.T. Issues and models concerning a finite number of inputs. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, N.J.: Erlbaum Assoc., 1974.

Trabasso, T., Rollins, H., & Shaugnessy, E. Storage and verification stages in processing concepts. *Cognitive Psychology*, 1971, *2*, 239-289.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Underwood, B.J. Individual differences as a crucible in theory construction. *American Psychologist*, 1975, *30*, 128-134.

Vernon, P.E. *The structure of human abilities*. London: Methaven, 1961.

von Arnin, G., & Engel, P. Mental retardation related to hypercalcaemia. *Develop. Med. Child Neurol.*, 1964, *6*, 366-377.

Warren, J.L. & Hung, E. Information processing in medically defined groups of retarded children. Paper presented at 9th annual Gatlinburg Conference on Research in Mental Retardation. Gatlinburg, Tenn., 1976.

Wason, P. & Johnson-Laird, P. *Psychology of reasoning*. Harvard: Harvard University Press, 1972.

Wertheimer, M. *Productive thinking*. New York: Harper & Row, 1959.

Wescourt, K.T. & Atkinson, R.C. Scanning for information in long and short-term memory. *Journal of Experimental Psychology*, 1973, *98*, 95-101.

Zellweger, H., & Schneider, J. Syndrome of hypotonia-hypomentia-hypogonadism-obesity (HHHO) or Prader-Willi syndrome. *Amer. J. Dis. Child.*, 1968, *115*, 588-598.

760801

Choice Reaction Time: What Role in Ability Measurement?

Clifford E. Lunneborg

University of Washington

Running head: Choice Reaction Time: What Role in Ability Measurement?

760801

Abstract

Three studies are described in which choice reaction time (RT) was related to such psychometric ability measures as verbal comprehension, numerical reasoning, hidden figures, and progressive matrices tests. Although fairly consistent negative correlations were found between these tests and choice RT when high school samples were used, differences from study to study highlight the need to develop more reliable measures for cognitive laboratory procedures and to study these in populations that are more broadly representative of human cognitive power.

760801

Choice Reaction Time: What Role in Ability Measurement?

Introduction

Two groups within psychology have been concerned with a common issue. Psychometricians have developed an array of individual difference constructs--verbal ability, perceptual speed, spatial visualization--to explain human intellectual performance. At the same time cognitive psychologists (e.g., Hunt 1971, 1973) have elaborated models of human information processing which include such constructs as rate of transfer of information, size of information store, and strategy availability. At issue in both of these developments is understanding how humans think.

Each approach has its strengths and through them can aid the other. Although the models developed by cognitive psychology provide challenging hypotheses about underlying processes, such models tend to describe the ideal human. Psychometrics is founded on the recognition that there is no ideal intelligence and has evolved the measurement of differences between people to the state of a science. By providing a methodology for building reliable measures of individual differences in the parameters of cognitive theories, psychometrics provides cognitive psychology with ways to test and modify their models. Such models can therefore become explanatory of the full range of human mental capacity. In return, cognitive psychology affords psychometrics a process orientation to intelligence which may be more promising than the present product orientation which equates intellectual abilities with achievements.

There should also be appreciable social gains to shifting the focus of individual differences from achievement to process. Intellectual assessment has come under considerable criticism for labeling as lacking ability or aptitude people who may simply not have had the opportunity achieve. A second advantage to such a shift would be to improve the diagnostic value of psychometrics, particularly in education. For example, the search for aptitude-treatment interactions as a way of improving education has been disappointing. The classical aptitudes have not interacted with different teaching styles to improve assignment of students to instructional techniques. However, if the information processing models identify "aptitudes" conceptually more closely tied to teaching strategies, then more efficient use of teaching and learning resources can be realized.

Psychologists have begun to study the relationships between the constructs of psychometrics and the processes of cognitive models. A series of laboratory studies at the University of Washington (UW) has provided information processing descriptions, for example, of high and low verbal subjects (Hunt, Lunneborg, and Lewis, 1975). Carroll (1974) has begun a similar line of research by describing the information processing requirements of each of the tests included in the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963). The present research was begun to relate Carroll's analytic efforts to the laboratory data being generated at UW.

Time to accomplish information processing tasks is a key element to the cognitive models and could be expected to show variability from subject to subject. Time is required, for example, to retrieve information from long term memory (LTM). Similarly, time is required to apprehend a stimulus. Both times can be expected to vary across persons. The extent of such variability must be established, and the covariation between such times as well, if we are to know which and how many parameters account for individual differences in intellectual performance. The simplest of the processing times, reaction time, is important because it is a component to all other processing times. Its variability and covariation with other measures need, therefore, to be determined first.

Depending on whether reaction time covaries with other information processing parameters, two predictions might be made about the correlations of it with paper and pencil test performance. If simple reaction time is independent of the other processing parameters, then Carroll's (1974) analysis would predict negative correlations with some tests and zero correlations with others. Negative correlations would result with tests where performance was largely determined by quickness in recognition and zero correlations would result with tests where performance was based on higher levels of processing. Negative correlations would result from shorter times to respond being associated with larger test scores. In contrast, the speculation of Hunt, Frost, and Lunneberg (1973) that individuals may have a characteristic "cycle" time suggests that all information processing times, including the simplest, are positively intercorrelated and would lead to negative correlations between reaction time and performance on all intellective tests. The present paper summarizes three studies designed to shed light on the role of simple choice reaction time in higher order intellectual functioning as well as to test out Carroll's hypothesized processing requirements for paper and pencil tests.

STUDY 1

The purpose of the first study was to examine the extent to which Carroll's (1974) description of the cognitive demands of several simple tests (French et al., 1963) was matched by correlations between performance on similar tests and laboratory measures of information processing. The study was based on a reanalysis of data collected prior to Carroll's paper so that not all of his cognitive tasks were included.

Method

Subjects

Sixty-four high school age participants, 33 men and 31 women, from the Seattle area were recruited during summer 1973 via a community Job Line. In groups of eight they spent half a day on each of two consecutive days in the laboratory. During each session they participated in several laboratory assessments and completed a battery of short paper and pencil tests. Participants were paid on an hourly basis.

Psychometric instruments

1. Employee Aptitude Survey (EAS) Verbal Comprehension (Ruch & Ruch, 1963) calls for the participant to select a synonym for each of a series of words. In the Reference Kit scheme this test measures Factor V, Verbal Comprehension.

2. EAS Numerical Ability requires participants to recognize the answers to simple arithmetic problems. This test measures Factor N, Number Facility.

3. EAS Space Visualization asks that adjacent blocks in a stack be counted. The test is considered a measure of Factor Vz, Visualization.

4. EAS Numerical Reasoning presents number series and asks that the next element be selected from among alternatives. This test taps Factor I, Induction, as well as Number Facility.

5. EAS Verbal Reasoning provides deductive reasoning tasks and is a measure of cognitive factor Rs, Syllogistic Reasoning. Proposed conclusions are reported to be true, false, or of indeterminate validity.

6. Clerical Number: This clerical test (Andrew & Paterson, 1946) requires the comparison of pairs of numbers and, subsequently, of pairs of personal or business names. Clerical Number measures Factor P, Perceptual Speed.

7. Clerical Name: Should also tap Factor P, Perceptual Speed, and is from same test as 6.

8. Hidden Figures from the French et al., (1963) Kit is a measure of Factor Cf, Flexibility of Closure. Participants are asked to identify which polygon is hidden (included) in a series of more complex figures.

Laboratory measures

1. Motor Reaction Time (RT): Median time to respond by key press to the onset of a + centered on a cathode ray tube (CRT) screen.

2. Choice Time: Median time to respond differentially by depressing

a left or right key to a circle either in the left or right half of the CRT display was calculated. Choice Time was the difference between this median time and median motor RT.

3. Prop Correct: Proportion of choice reaction time trials on which the correct choice was made.

4. Stroop Difference: In a modification of the Stroop task participants were asked first to name the colors in which a series of asterisks were printed and then to name the color of printing of an equivalent length list of contrasting color names. The task was repeated twice and the Stroop Difference score was the average difference in "reading times" between the name and asterisk conditions.

5. Search Slope: In an experiment in the Sternberg (1970) paradigm, one to six consonants were sequentially shown on a projection screen followed by a single probe consonant. Participants were asked whether the probe was in the previously exposed set. Response times were recorded and, for correctly identified instances, these times linearly regressed on the number of digits in the associated target set. Search Slope was the slope of this regression line.

6. Search Intercept: Intercept of the above regression line.

7. Digit Span Final: Strings of fifteen digits were presented binaurally with recall cued immediately following presentation. Recall instructions were for subjects to report in order as many digits as possible beginning with the first digit heard. Ten trials were presented with each scored for number of digits recalled in order. Digit Span Final is the average of performance on the last five trials.

8. Digit Span Gain: Digit Span Final minus average performance on the first five trials.

9. Dichotic Category: In a task patterned after Massaro (1972) four digits and four consonants were presented dichotically, two digits and two consonants to each ear. Following presentation recall was requested. On some blocks of trials participants were to report by ear and on others by category (digit or consonant). The Dichotic Category score was the number of items correctly reported over all 40 category trials.

10. Dichotic Difference: Score based on ear trials alone minus Dichotic Category score.

11. Clustering Base: A final set of four scores was obtained from a semantic clustering task based on the experimental paradigm of Puff (1966). During their first laboratory session students were shown item by item, two lists of 30 common nouns. Each list consisted of ten nouns from each of three semantic categories (fruits, occupations, animals etc.). Each list was shown twice with recall asked for immediately following each presentation (four recalls, first day). For half the students the list presented first was blocked, all ten members of a given category appearing contiguously, and the list presented second was in pseudo-random order. For the other half of the sample this order was reversed. When participants returned for the second day in the laboratory they were asked to recall each of the two lists (two additional recalls, second day). A semantic clustering score was computed for each of the

six recalled lists. Clustering Base was the clustering score for the second presentation of the blocked list (first day).

12. Clustering Difference: Clustering Base minus clustering score for second presentation of pseudo-random list, also on first day.

13. Clustering Delay: Clustering Base minus clustering score for second day recall of the blocked list.

14. Clustering Delay Difference: Clustering score for second day recall of blocked list minus clustering score for second day recall of pseudo-random list.

There are above three types of laboratory response measures: time (Motor RT, Choice Time, Stroop Difference, Search Slope, Search Intercept), response accuracy (Prop Correct, Digit Span Final, Digit Span Gain, Dichotic Category, Dichotic Difference) and response clustering (Clustering Base, Difference, Delay, and Delay Difference).

Results

Correlations between the psychometric and laboratory data are shown in Table 1. The five response time measures correlated negatively with the paper and pencil test performances. As for the accuracy measures, Prop Correct correlated weakly, though positively, with the psychometric measures. Prop Correct, however, had little variability as few errors were made (average proportion correct was .93). The negative correlations involving Digit Span Gain suggest that inasmuch as average improvement was small (from 5.75 digits to 6.23 digits), big gains were registered by students who had underperformed on the initial trials. Digit Span Final scores were not highly correlated with any of the psychometric measures.

[Insert Table 1 about here]

The Dichotic Category score correlated positively with nearly all tests. Dichotic Difference (ear minus category scores), on the other hand, correlated negatively, suggesting that high test scores were earned by students with category scores much higher than ear scores. (For nearly everyone the ear scores were smaller than the category scores.)

The clustering score , based on the number of interruptions of semantic strings, was such that lower scores indicated greater semantic clustering. Except for the reversal of signs, Clustering Base showed the same strength of correlation as the Dichotic Category scores. Clustering Delay showed some negative correlation whereas Clustering Difference and Clustering Delay Difference were not correlated with the psychometric tasks at all. Remembering that Clustering Delay represents first day recall minus second day recall, high scores are associated with increased reliance on semantic clustering between immediate and delayed recall. Thus, increasing reliance upon clustering was negatively related to performance on the reasoning tests and on Hidden Figures.

Table 2 illustrates how the laboratory measures were correlated among themselves. Response time measures were positively intercorrelated. Accuracy measures, however, were not consistent. Prop Correct, Digit Span Final, and Dichotic Category were not all positively intercorrelated, nor were the three consistently negatively related to Clustering Base. Clustering Base and Dichotic Category were strongly related ($\underline{r} = -.55$) suggesting that both provide measures of a common accuracy of immediate recall. Although the Clustering Base score is a measure of semantic clustering it is derived from immediate recall of a blocked list. Hence, accurate serial recall of the list would yield a high degree of semantic clustering. Lastly, correlations of Motor RT, Choice Time, Search Slope, and Search Intercept with Dichotic Category and Clustering Base indicate that speed and accuracy of response are related.

[Insert Table 2 about here]

Because of the overlap among the laboratory measures a series of multiple regression analyses was conducted. Each of the psychometric tests was taken as a criterion measure with Motor RT and Choice Time forced as the first two predictors. Additional laboratory predictors were selected so long as they significantly ($\underline{p} < .05$) increased $\underline{R}^2$. This strategy was adopted to evaluate first what proportion of test variance could be accounted for by the two simplest reaction time measures, and second, to determine which if any higher order information processing measures contributed as well. While most researchers would agree that the first two measures require the least intellectual activity, it was not possible to arrange the remaining measures in terms of increasing demands upon the information processer. The results of predictor selections are given in Table 3.

[Insert Table 3 about here]

Motor RT was significantly related to all but one of the tests, Clerical Number. Choice Time added significant variance to five out of the eight criteria, the exceptions being Verbal Reasoning, Clerical Number, and Clerical Name. Together the variance accounted for by these two simple reaction times ranged from only 8% for Clerical Number to 34% for Space Visualization. For five of the eight measures the first freely chosen predictor was Clustering Base and in four of these five instances it terminated selection. In these four cases the variance shared with the laboratory measures rested upon speed of response and accuracy of short term recall. This variance ranged from 25% for Verbal Reasoning to 49% for Numerical Reasoning. The third and only added predictor for Clerical Number and Clerical Name was Stroop Difference. For these two perceptual speed tests, Stroop Difference also had the highest zero order correlations (Table 1). Stroop Difference thus adds to the time required for a simple comparison (Choice Time), that required to look up a name. The two most complex tests in terms of the number of laboratory measures providing increments were Verbal Comprehension (vocabulary) and Hidden Figures (field independence). For both Dichotic Category and Clustering Delay Difference were selected. Overall,the proportion of total test variance accounted for by laboratory measures ranged from 16% to 51%. The upper figure is quite high given the reliability of such short tests.

Three aspects to these results are worth emphasizing. First, Motor RT was significantly correlated with all but one of the psychometric tests. Thus, the time it takes to accomplish the simplest processing task seems to contribute more to intellectual skills than might offhand be expected. Second, Verbal Comprehension had the bulk of its reliable variance determined by these laboratory measures. This occurred despite the fact that none of the laboratory measures assessed the size of memory, e.g., extent of vocabulary. Finally, despite the variety of laboratory timed measures, surprisingly little of the variance on Clerical Number and Clerical Name was accounted for. Such perceptual speed tests should have had more in common with response time measures.

STUDY 2

Certain deficits to Study 1 were considered in designing Study 2 which, again,sought to increase basic understanding of the relationships between the intellective factors of psychometrics and the processes of cognitive models. All of the psychometric tests in Study 1 were quite short. This may have contributed to their strong correlations with reaction time by placing premium on speed of answering test items. Secondly, each intellective factor, except Perceptual Speed, was represented by only a single test. Thirdly, many important cognitive factors were not represented at all, and finally, only single, summary motor RT and choice time scores were available, not permitting study of the differences between an initial and a learned rate of response.

Method

Subjects

First year University students were recruited fall 1974 from those who had completed the battery of tests administered high school juniors by the Washington Pre-College (WPC) Testing Program (WPC, 1975). Volunteers were selected so that the joint distribution of the WPC Verbal and Quantitative Composite scores in the sample reflected the distribution for the entire entering class. Volunteers were paid and their participation extended over an academic quarter. Each student was involved in one psychometric and one laboratory session each week. The total group consisted of 27 men and 36 women.

Psychometric instruments

Because of the large number of tests involved, all were first intercorrelated and then separate analyses by type undertaken. The resulting reduced set of measures was finally related to laboratory performance.

Verbal ability measures

The WPC battery provided scores of Reading Comprehension, Vocabulary, Spelling, and English Usage (Grammar). (The Verbal Composite score used in subject selection was a linear weighting of these four tests.) Other verbal tests, administered following recruitment, included EAS Verbal Comprehension from Study 1 and the two subparts of the Terman Concept Mastery Test (Terman, 1956), Synonyms and Antonyms and Verbal Analogies. In the French et al. (1963) scheme all of these tests would be considered measures of Factor V: Verbal Comprehension.

A principal components analysis of the correlations among these verbal measures followed by a varimax rotation of the larger of these components resulted in a reduction to two measures:

1. Vocabulary. Average of standardized WPC Vocabulary and Terman Synonym-Antonym scores.
2. Grammar. Average of standardized WPC Spelling and English Usage scores.

Quantitative ability measures. The WPC battery provided three quantitative tests: Quantitative Skills (numeric reasoning), Applied

Mathematics (word problems at the elementary algebra level) and Mathematics Achievement (achievement in high school mathematics). (The three were linearly combined to provide the Quantitative Composite measure used in subject selection.) Additional quantitative tests administered this sample included EAS Numerical Ability and Numerical Reasoning described in Study 1. A principal components analysis was conducted and one component appeared to tap the variability central to all these five measures, Number Facility (N). A separate Reasoning (R) factor was not found and, thus only a single quantitative measure was retained:

3. Quantitative Composite. The WPC Quantitative Composite score.

Spatial ability measures. One spatial ability measure was included in the WPC battery, a paper folding test of the kind classified by French et al. (1963) as Visualization (Vz). Two other visualization measures were administered, the Two Dimensional and Three Dimensional Spatial Relations tests of the California Multiple Aptitudes Test (CMAT) (Segel & Raskin, 1959). The two dimensional test is a paper form board and the three dimensional test one of surface development.

Other figural tests included Hidden Figures from Study 1 which loads the Flexibility of Closure (Cf) factor (and, to a lesser degree, the Spatial Orientation , S, factor), a Maze Tracing task (French et al., 1963) associated with a Spatial Scanning (Ss) factor, and the Advanced Progressive Matrices (Set II) of Raven (1962). The Progressive Matrices test is variously considered a measure of general intelligence, an indicator of a reasoning factor, Induction (I), or a "space" measure. It was included in this set because, in the present study, it had sizeable correlations with the other figural measures.

The principal components/varimax rotation analyses of these spatial tests defined two clusters and two new composite measures were calculated.

4. Space I. Average of standardized Maze Tracing and paper form board test scores.
5. Space II. Average of standardized Progressive Matrices, Hidden Figures, surface development, and paper folding scores.

Reasoning ability tests. All of the tests in this category were administered after subject recruitment. EAS tests of Verbal Reasoning (from Study 1) and Symbolic Reasoning (involves the concepts of equality and inequality presented symbolically) tapped Syllogistic Reasoning (Rs). The Watson-Glaser Critical Thinking Analysis (Watson & Glaser, 1964) provided scores on five subtests: Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. With the possible exception of Deduction these subtests are probably factorially complex, requiring a comprehension of material read and its integration with some already known information as well as the exercise of reasoning.

When the interrelations among all tests were examined, Symbolic Reasoning was more strongly related to quantitative measures, while EAS Verbal Reasoning and Watson-Glaser Recognition of Assumptions, Deduction, and Interpretation, were moderately correlated with the verbal measures. In consequence, the two reasoning measures used here were those independent of other factors:

6. Inference. After reading a paragraph a series of statements are classified as true, probably true, false, probably false or based on insufficient data.
7. Evaluation of Arguments. A "policy question" is followed by a number of pro and con statements to be classified as weak or strong arguments on the question.

Perceptual/motor speed tests. These tests, again all administered after subject recruitment, placed a premium on quick perception and response. The survey of Working Speed and Accuracy (Ruch, 1943) included Number Checking (comparing pairs of numbers for identity), Code Translation (substituting letters for integers), Finger Dexterity (dotting small circles), and Vowel Counting (obtaining counts for lines of test). Number Checking, Vowel Counting, and possibly Code Translation are measures of the Perceptual Speed (P) factor, whereas Finger Dexterity has been regarded as motor rather than intellective.

A Code Recognition test was also administered from the Flanagan Aptitude Classification (Flanagan, 1953) battery. Following practice in learning alphanumeric codes for common objects, participants were tested immediately on recognition of these codes. The test is a recognition form of the measures (French et al., 1963) used as markers for the Associative Memory (Ma) factor.

Finally, three locally constructed tests were administered, based upon descriptions given by Rose (1974) of a battery developed at the Human Performance Center, University of Michigan. The three were a Grammatical Reasoning test (requiring comparison of a statement "A does not precede B," with a letter pair AB or BA), a Letter Search test (checking blocks of "randomly" assembled letters for a key letter or letters), and a Tapping test in which circular targets of varying size and separation are to be dotted.

Letter Search should be another Perceptual Speed test. Grammatical Reasoning was used as a potential paper and pencil surrogate for a laboratory procedure used in earlier studies (Hunt, Lunneborg & Lewis, 1975) and based on the work of Clark and Chase (1972). The tapping task was intended to provide a paper and pencil measure of a fairly unencumbered response speed. Because these three tests were so simple they were administered under highly speeded conditions on each of two occasions separated by about three weeks. Upon examination, the between sessions correlations for the same test were disappointingly low, while the between tests correlations for the same, final session were all in excess of .90. As a representative of this set of tests, one score was retained:

8. Letter Search. Average of standardized scores for the searches for 1, 2, and 4 letters on the final administration.

A principal components analysis of the correlations among the other perceptual/motor scores yielded three dimensions represented in turn by Number Checking, Code Translation, and Vowel Counting. As Vowel Counting was strongly correlated with already selected Letter Search only the first two measures were retained:

9. Number Checking.
10. Code Translation.

Of the ten psychometric measures, six were based on essentially unspeeded performance, while four, Letter Search, Number Checking, Code Translation, and Space I, were more speed dependent.

Laboratory measures

As with the psychometric measures, time to respond data from several laboratory procedures were first reduced by a series of principal components analyses. The resulting set of measures was then related to the ten psychometric measures described above.

Binary choice reaction time. Participants fixated a midpoint on a CRT display and responded by key press to the onset of a character to the left or right of that midpoint. Each student participated under two conditions. In the one-finger condition, the index finger of the preferred hand was used to depress one of two keys. In the two-finger condition keys under left or right hand index fingers were used to signal detection of display to left or right respectively. Data were median response times, separately for presentions to the left and right, for the first and last (sixth) blocks of 100 presentations under each of the conditions. Principal components analysis yielded two components and as a result two composite measures were used:

1. Choice Reaction Time One Finger. Average of standardized scores for left and right presentations, first and last block.
2. Choice Reaction Time Two Finger. Same average based on two-finger condition.

Delayed auditory feedback times. In this procedure participants read aloud a page length stream of nonsense syllables under three conditions: reading was taped and played back through earphones at zero, 150, and 300 milliseconds delay. Although types of errors actually motivated this task, two time scores were of interest here:

3. No Delay Reading Time. Time to read the passage.
4. Delay Reading Time. Average of two standardized times to complete reading when feedback was delayed.

During this same experimental session participants repeated to the experimenter a series of words replacing R's with L's and L's with R's. The score obtained was:

5. R/L Latency. Median time to transform a word.

Raven response times. Set I of the Advanced Progressive Matrices was presented individually, one problem at a time, and solution time recorded for each figure. This was done after group administration of Set II and involved 12 problems not in the larger set. Principal components analysis of these time scores produced two components based respectively on earlier and later problems in the set. Two scores were thus developed:

6. Raven Time I. Average of standardized times to solve problems 1 and 2.
7. Raven Time II. Average of standardized times to solve problems 8 and 9.

Verbal problem solving time. Another laboratory task called for learning a list of statements which associated each of a set of personal names either with occupations (Ruth is a lawyer) or locations (John is in the pool), or which associated locations with activities (If you are in the park you can hear the chimes ring). A week later following a recall test of mastery, a series of True or False questions was presented. The questions were such as to require the retrieval of 1, 2, 3, or 4 of the learned facts. Time to answer each was recorded and two scores of interest derived:

8. Verbal Problem Solving Slope. For the items answered correctly time to respond was linearly regressed against number of facts required. The problem-solving slope is the slope of this line.
9. Verbal Problem Solving Intercept. Intercept of line in 8.

Results

Table 4 reports zero order correlations between the psychometric and laboratory measures. Choice reaction times were not significantly related to the psychometric measures, not even to the speeded measures. This result contrasts sharply with the results in Table 1 for the high school group of Study 1. Reading time with delayed auditory feedback, time to exchange R's and L's, the intercept of the problem solving time line, and more weakly, solution time for Raven problems 1 and 2 yielded correlations that most closely resembled the expected pattern of negative relations. But, surprisingly, problem solving slope was positively related to certain (unspeeded) tests. Because the results of the two studies differed so greatly, the stepwise regression analyses of Study 1 were not repeated. Rather, principal components and canonical correlation analyses were done to explore the relations between psychometric performance and laboratory measures in this instance. Would even greater combining of measures into components reveal relationships more like those of Study 1?

[Insert Table 4 about here]

A principal components analysis and varimax rotation of the nine laboratory measures yielded the four components summarized in Table 5. Component 1 was loaded by time to respond measures when the processing load was minimal. Component 3, in contrast, reflects time to respond with more pressure on the system (delay of auditory feedback, overcoming phonetic convention, searching LTM for learned facts). Component 4 associated time to read through an LTM store (Problem Solving Slope) with time to read through a page of nonsense syllables (No Delay Read Time). Time to solve Raven problems appears independent of the time to solve verbal problems, defining a separate, second component. The Raven items differ from the verbal problems in that they are symbolic rather than verbal, become quite difficult, and depend for solution on more than recall of a closed set of recently acquired facts.

[Insert Table 5 about here]

Table 6 reports the results of a canonical correlation analysis involving the psychometric and laboratory measures. Two significant correlations were obtained: $\underline{r} = .86$ for the first pair of canonical variates and $\underline{r} = .74$ for the second pair. The first pair involves large positive weights for the following response time measures: Choice RT Two Finger, Delay Read

Time, Verbal Problem Solving Slope and Intercept. On the psychometric side positive weights are assigned the Quantitative Composite and unspeeded Space II scores, and negative weights for the Grammar and speeded Space I scores. The second pair weights R/L Latency positively and Verbal Problem Solving Slope negatively on the laboratory side, while on the psychometric side weighting speeded Space I positively and Vocabulary negatively.

[Insert Table 6 about here]

These canonical variates are not easily interpreted. For both a negative relationship between Space I and the Verbal Problem Solving Slope contributes. As a result it is difficult to use this relationship to characterize either variate. Leaving it aside, the first pair of variates could be seen in terms of Grammar and Verbal Problem Solving Intercept as tapping something like speed of rule application. In both instances subjects consult a set of overlearned rules and determine whether they apply in a particular situation. The faster this is accomplished in the problem solving task (low Intercept), the higher the Grammar score. Similarly, ignoring Space I and Verbal Problem Solving Slope, the interpretation of the second variate depends on Vocabulary and R/L Latency. This suggests the second variate measures verbal fluency: slowness in converting R's to L's is associated with poor vocabulary. At best, however, these characterizations are only suggestive of the underlying bridge between intellectual performance and simple response times. Unfortunately, none of the components identified in Table 5 appeared as part of either canonical variate. Thus, whatever linked the laboratory measures with one another did not serve to relate them in turn to paper and pencil test performance.

STUDY 3

The failure in Study 2 of the choice reaction time measures to show stronger relations with the psychometric measures, in particular the failure of these laboratory scores to load together with the more speeded paper and pencil tests in a single variate, was unexpected given the results of Study 1. The differences in findings could have resulted from differences in measurement techniques, from differences in subject sampling, from both or from other, less obvious differences between the two studies. The most parsimonious follow-up, however, was the administration of Study 2 measures to a sample of subjects like those of Study 1. Would the same strong relations be found if students at the high school level were the sample under investigation?

Method

Subjects

At the end of spring term 1975 high school juniors who had completed the WPC battery were recruited through their counselors. Sixty-four volunteers, 30 women and 34 men, were run in groups of eight for two days. Three hour morning sessions were devoted to psychomtric testing and slightly longer afternoon sessions involved laboratory data collection. As in the earlier studies, participants were paid at an hourly student rate.

Measures

Although Study 3 involved other laboratory tasks, results will be reported here only on that subset which is in common with Study 1 and, to a greater extent, Study 2. In particular, scores were obtained on ten psychometric instruments described earlier:

1. EAS Verbal Comprehension
2. EAS Numerical Ability
3. EAS Numerical Reasoning
4. EAS Verbal Reasoning
5. EAS Symbolic Reasoning
6. Advanced Progressive Matrices (Set II)
7. Maze Tracing
8. Hidden Figures
9. Two Dimensional Spatial Relations (CMAT)
10. Three Dimensional Spatial Relations (CMAT)

Tests were administered under all of the same conditions as for Study 2 and, where applicable, Study 1.

The Binary Choice Reaction Time task of Study 2 was repeated in a slightly modified format. Inspection of individual records in Study 2 suggested that stabilization of response was as likely after 300 trials as after 600. In Study 3, therefore, only 300 trials (exclusive of an initial block of 30 warm-up trials) were completed. Only the faster Choice Reaction Time Two Finger time was employed here.

Results

Table 7 summarizes the zero order correlations between the binary choice reaction time tasks and sets of psychometric scores for Studies 1 (high school 1973) and 2 (college 1975) as well as Study 3 (high school 1975). Data from the first two studies are reported in Table 7 in slightly different form than in previous tables: actual choice reaction times from Study 1 are used rather than "Choice Time" (the difference between choice and motor reaction times) and, for Study 2, choice reaction times are reported separately for the first and last (sixth) blocks of 100 trials.

[Insert Table 7 about here]

The significant relations of Study 1 were not replicated for either the one finger or two finger tasks with the college sample. For the second high school group, however, there was a tendency for this negative relation to emerge again for Block 3 data. (In connection with Tables 7, 8, and 9 it should be noted that the spatial visualization test used in Study 1 is different from those tests administered in Studies 2 and 3. Data for the Study 1 test are thus enclosed in parentheses.)

Means and standard deviations for the three groups on these variables are reported in Table 8. In terms of average psychometric performance, the Study 3 high school group was only slightly less proficient than the college group. The Study 1 high school group, however, was markedly less able, e.g., a mean of 15.8 on Numerical Ability as compared to 26.2 for the Study 3 high school students. Choice reaction times, in contrast, were not markedly different among the three groups. Of greater significance than the differences among these means are the differences in the standard deviations. While there was no consistent pattern of differences in the S.D.s on the psychometric instruments, the two high school groups were appreciably more variable on the choice reaction time tasks than the college group.

[Insert Table 8 about here]

Variability is known to affect correlations greatly. Could the reduced correlations in Table 7 for the college students have resulted from their smaller RT standard deviations? This possibility can be evaluated by converting the correlations to regression coefficients. While a correlation coefficient indicates the strength of a relation in terms of the proportion of variance in one measure accounted for by a second--and is, hence, sensitive to the amount of variance present--a regression coefficient measures the same relation in terms of the amount of change in one measure that would be expected to accompany a fixed amount of change in a second.

Will the relations between Choice RT and psychometric performance appear more consistent if regression coefficients replace correlations? The regression coefficients will have the same signs as the corresponding correlations and any inconsistencies in direction of relationship will not be corrected. Thus, the positive correlations in Table 7 will remain unresolved. (None of these positive correlations, however, were significantly different from zero.) Regression coefficients were computed for the negative correlations in Table 7.

[Insert Table 9 about here]

To make the regression coefficients of Table 9 a little easier to understand the coefficients have been scaled: (1) Negative signs have been changed to positive so that the entries now express amount by which test scores would be expected to decrease with increases in Choice RT. (Remember as reaction times got larger (slower), test scores decreased in Study 1.) (2) Because the unit of measurement for Choice RT was so small, a millisecond, each regression coefficient has also been multiplied by ten. The result of the two scaling changes is that the coefficients in Table 9 are interpreted as the amount by which a test score may be expected to decrease as a result of increasing the Choice RT by one hundredth of a second (10 ms.).

The reanalysis presented in Table 9 only partly clarifies the relationship of choice RT to paper and pencil test performance. Wholly consistent results would have been represented by a constant value for any given row (test). What resulted instead can be summarized as follows: (1) The strength of relationship between choice RT and test performance was still higher for high school students than for college students even after adjusting in this way for differences in variability. (2) The two high school studies provided different results. The EAS tests which were strongly correlated with choice RT in Study 1 were not the tests which were strongly correlated with choice RT in Study 3. Spatial tests had the bulk of the strong correlations in Study 3. (3) Which choice RT to relate to psychometric performance remains in question. For the Study 3 group it was performance in the third block of 100 trials, while for Study 2, using the same Choice RT Two Finger task, initial performance (first block of 100 trials) was more consistent.

Discussion

The implications of these three studies would seem to be that there is no simply obtained answer to the question of how choice RT relates to paper and pencil test performance. The patterns of correlations between these two types of intellective measures were different among the three samples. The college sample showed the weakest relationships which may be attributable to their lower variability on the choice RT tasks. This group would appear to be so selective as to be a questionable population in which to explore such relationships. The differences in correlations between the high school groups, on the other hand, cannot be explained on this basis. They shared the same range of choice RT performance. They did differ, however, in absolute level of intellective ability, the Study 3 group being much more capable than the Study 1 group. Curiously enough, while the strong correlations of Study 1 were with verbal tests, the strong correlations of Study 3 were with spatial tests.

Some part of the observed differences in the patterns of correlations are likely to also be attributable to changes in the technique of measuring choice RT. Use of one or two handed tasks and basing the measure on an earlier or later blocks of trials produced different results. These two aspects must be regarded as part of the definition of the choice RT to be studied. Equally important to this, the selection of a choice RT measurement technique can also be expected to influence the reliability (within subjects consistency) of those observations.

How should one investigate the role of choice RT in ability measurement in light of these results? First, great care must be given to the definition of the subject population. Characteristics desired in such a population are a level and range of performance on both sides, cognitive and psychometric, which are no more restrictive than what would be found in sophomore or junior high school students. Larger samples than were used here will necessarily be required to meet this goal.

Second, equally great care must be given to the definition of the choice RT task. Here this should be directed towards obtaining the most reliable assessment. This quality seems more desirable than the theoretical constraints of being "initial" or "practiced." The typical subject in the laboratory performs erratically at first, then stabilizes, and finally grows erratic again as boredom and fatigue overpower the intrinsic motivation provided by the task. This probably takes place at different rates for different people. Given this situation, a tailored testing approach to choice RT measurement seems worthy of exploration, i.e., determining for each person a choice RT for that most stable sequence of trials wherever it occurs in the longer experimental paradigm. That median RT for left and right hand responses during the terminal, sixth block of 100 trials correlated only .60 in Study 2 provides testimony on this point.

Despite the inconclusive results of studies such as these, the growth of cognitive psychology coupled with disillusionment with the apparent shortcomings of classical psychometric assessment is certain to result in continuing attempts to develop measures of individual differences in intellective ability based on cognitive theory and laboratory work. These attempts will be the more effective if the two prime dicta of individual differences psychology, just exemplified, are kept in mind: (1) Defining relevant normative populations, and (2) Developing reliable assessments.

References

Andrew, D. M., & Paterson, D. G. Manual for the Minnesota vocational test for clerical workers. New York: The Psychological Corporation, 1946.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." Paper presented at the LDRC Conference on the Nature of Intelligence, Pittsburgh, March 1974.

Clark, H., & Chase, W. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Flanagan, J. C. Flanagan aptitude classification test. Chicago: Science Research Associates, 1953.

French, J. W., Ekstrom, R. B., & Price, L. A. Manual for kit of reference tests for cognitive factors. (Revised 1963) Princeton, N. J.: Educational Testing Service, 1963.

Hunt, E. B. What kind of computer is man? Cognitive Psychology, 1971, 2, 57-98.

Hunt, E. B. The memory we must have. In R. Schank & K. Colby (Eds.), Computer simulation of information processes in man. New York: W. H. Freeman, 1973.

Hunt, E. B., Frost, N., & Lunneborg, C. E. Individual differences in cognition. In G. Bower (Ed.), Advances in learning and memory, Vol. 7. New York: Academic Press, 1973.

Hunt, E. B., Lunneborg, C. E., & Lewis, J. What does it mean to be high verbal. Cognitive Psychology, 1975, 7, 194-227.

Massaro, D. W. Preperceptual and synthesized auditory storage. Technical report 72-1. Madison: University of Wisconsin Psychology Department, 1972.

Puff, C. R. Clustering as a function of sequential organization of stimulus word lists. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 503-506.

Raven, J. C. Advanced progressive matrices (1962). London: H. K. Lewis, 1962.

Rose, A. M. Human information processing: an assessment and research battery. Ann Arbor: Human Performance Center, Department of Psychology, University of Michigan. Technical Report 46, 1974.

Ruch, F. L. Manual for survey of working speed and accuracy. Los Angeles: California Test Bureau, 1943.

Ruch, F. L., & Ruch, W. W. Employee aptitude survey technical report. Los Angeles: Psychological Services, Inc., 1963.

Segel, D., & Raskin, E. Multiple aptitude tests manual. Los Angeles: California Test Bureau, 1959.

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. In J. S. Antrobus (Ed.), Cognition and affect. Boston: Little-Brown, 1970.

Terman, L. M. Concept mastery test manual. New York: The Psychological Corporation, 1956.

Washington Pre-College Testing Program Learn more about yourself--explore your future: guide to the Washington pre-college test. Seattle: Washington Pre-College Testing Program, 1975.

Watson, G., & Glaser, E. M. _Manual for the Watson-Glaser critical thinking appraisal_. New York: Harcourt-Brace & World, 1964.

Acknowledgements

This research was sponsored by the National Institute of Mental Health (grant MH21795, Earl B. Hunt, principal investigator) and by the National Institute of Education (grant NIE-G-74-0104, Clifford E. Lunneborg, principal investigator). Portions of this report were presented to the annual conventions of the Americal Psychological Association in 1974 and 1975. The assistance to the principal investigators of Kenneth Burkhardt, Jon Palmer and Steven Poltrock in the design and conduct of the laboratory studies is gratefully acknowledged.

Address of author: Clifford E. Lunneborg
Department of Psychology NI-25
University of Washington
Seattle, WA 98195

Table 1

Correlations between Psychometric and Laboratory Measures, Study 1[a]

Laboratory measure	Verbal comprehension	Numerical ability	Space visualiza	Numerical reasoning	Verbal reasoning	Clerical number	Clerical name	Hidden figures
Motor RT	-40**	-38**	-42**	-40**	-27*	-17	-38**	-27*
Choice time	-40**	-45**	-55**	-50**	-30*	-28*	-37**	-40**
Prop correct	21	11	12	04	14	10	14	17
Stroop difference	04	-27*	-19	-32**	-16	-35**	-38**	-09
Search slope	-07	-10	-27*	-16	-07	-24	-26*	03
Search intercept	-28*	-36**	-49**	-31*	-16	-29*	-39**	-27*
Digit span final	-04	02	-03	02	-12	22	16	05
Digit span gain	-10	-26*	-24	-21	-29*	-06	-17	-07
Dichotic category	55**	46**	49**	52**	32**	02	26*	48**
Dichotic difference	-21	-29*	-23	-31*	-18	-16	-29*	-25*
Clustering base	-52**	-48**	-48**	-61**	-47**	-02	-37**	-34**
Clustering difference	06	10	10	-07	07	00	09	-14
Clustering delay	-12	-22	-21	-28*	-25*	01	-03	-35**
Clustering delay dif	-18	00	06	04	05	06	00	20

[a] Decimal points omitted.

*$p < .05$
**$p < .01$

Table 2

Intercorrelations among Laboratory Measures, Study 1[a]

	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Motor RT	47**	-07	21	42**	69**	00	15	-52**	38**	36**	-28*	09	-19
2. Choice time		14	27*	43**	55**	-21	-11	-40**	30*	30*	03	04	07
3. Prop correct			14	25*	12	-29*	-41**	08	06	-20	17	-19	16
4. Stroop difference				27*	21	-32**	16	-11	22	14	-02	-07	-15
5. Search slope					39**	-17	-11	-21	17	15	-44**	-21	05
6. Search intercept						-21	-07	-45**	44**	21	-18	-02	-06
7. Digit span final							65**	-04	-02	14	-19	24	-28*
8. Digit span gain								-12	02	34**	-17	32**	-25*
9. Dichotic category									-75**	-55**	04	-35**	02
10. Dichotic difference										32**	08	26*	02
11. Clustering base											16	54**	09
12. Clustering difference												28*	32**
13. Clustering delay													-32**
14. Clustering delay dif													

[a] Decimal points omitted.

*$p < .05$
**$p < .01$

Table 3

Psychometric Performance Predicted from Laboratory Measures: Increments to the Squared Multiple Correlation for Selected Predictors[a]

Verbal comprehension			Numerical ability			Space visualiza			Numerical reasoning		
Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ
Motor RT	16	16	Motor RT	14	14	Motor RT	18	18	Motor RT	16	16
Choice time	22	06	Choice time	24	10	Choice time	34	16	Choice time	28	12
Clus base	35	13	Clus base	34	10	Clus base	42	08	Clus base	49	20
Dichot cat	40	05									
Dichot dif	47	07									
Clus deldif	51	04									

Verbal reasoning			Clerical number			Clerical name			Hidden figures		
Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ
Motor RT	07	07	Motor RT	03	03[b]	Motor RT	14	14	Motor RT	07	07
Choice time	11	04[b]	Choice time	08	05[b]	Choice time	19	05[b]	Choice time	17	10
Clus base	25	14	Stroop dif	16	08	Stroop dif	26	07	Dichot cat	28	12
									Search slp	34	05
									Clus deldif	38	04

---

[a] Decimal points omitted. Motor RT and Choice time forced as first and second selections; additional predictors selected only where R^2 significantly increased ($p < .05$). Δ is size of increment to R^2.

[b] Motor RT or Choice time not significant.

Table 4

Correlations between Laboratory and Psychometric Measures, Study 2[a]

	Psychometric measures									
Laboratory measures	Vocab	Grammar	Quant comp	Space I(S)	Space II	Inference	Eval argument	Letter search (S)	Number check (S)	Code trans (S)
Choice RT 1 finger	10	-17	-16	-13	05	-09	-21	-06	02	-11
Choice RT 2 finger	-10	-19	-14	01	-03	00	-21	07	11	00
No delay read time	03	-10	14	00	17	-02	04	-08	11	-05
Delay read time	-31*	-45**	-22	04	12	-22	-38**	10	-11	-21
R/L latency	-54**	-46**	-38**	-05	-12	-11	-07	11	01	-09
Raven time I	-12	-18	-22	-30*	-36**	05	09	12	-10	14
Raven time II	-04	-07	02	-13	-08	01	09	10	06	-05
Verbal prob solv slope	28*	01	23	-36**	03	25*	-05	-05	-07	-02
Verbal prob solv inter	-16	-45**	-11	02	17	-26*	-06	16	-44**	-26*

[a] Decimal points omitted. (S) denotes speeded psychometric scores.

$*p < .05$
$**p < .01$

Table 5

Loadings of Laboratory Time Measures on Varimax Rotated Principal Components, Study 2[a]

Measure	Component 1	2	3	4
Choice RT 1 finger	71			
Choice RT 2 finger	93			
No delay read time	42			64
Delay read time			83	
R/L latency			68	
Raven time I		89		
Raven time II		81		
Verbal prob solv slope				81
Verbal prob solv inter			67	

[a] Only correlations of .40 or larger indicated. Decimal points omitted.

Table 6

Canonical Correlations between Psychometric and Laboratory Measures[a]

Psychometric measures	Variate 1 weights	Variate 2 weights	Laboratory measures	Variate 1 weights	Variate 2 weights
Vocab	-01	-66	Choice RT 1 finger	-34	-06
Grammar	-99	-22	Choice RT 2 finger	56	-03
Quant comp	49	-21	No delay read time	00	11
Space I (S)	-51	58	Delay read time	53	-04
Space II	53	-09	R/L latency	-08	70
Inference	02	-19	Raven time I	02	10
Eval argument	-22	23	Raven time II	23	06
Letter search (S)	33	-08	Verbal prob solv slope	52	-76
Number check (S)	-09	22	Verbal prob solv inter	68	05
Code trans (S)	-07	15			

[a] Decimal points omitted. The first canonical correlation was .86 and the second .74.

Table 7

Correlations between Laboratory Choice Reaction Times and Psychometric Performances, Studies 1, 2 and 3[a]

	Choice RT one finger			Choice RT two finger			
	High school 1973	College 1975		College 1975		High school 1975	
		Block 1	Block 6	Block 1	Block 6	Block 1	Block 3
Verbal comprehension	-46**	10	14	-11	04	14	09
Numerical ability	-48**	-06	-02	-19	04	-04	-19
Numerical reasoning	-51**	-10	-08	-22*	04	-15	-12
Verbal reasoning	-33**	-04	-24*	02	17	-00	05
Symbolic reasoning		-08	-17	-08	05	-08	-31**
Progressive matrices		-05	-00	-11	-02	11	-27*
Maze tracing		-12	-19	-03	00	-29**	-28*
Hidden figures	-38**	-03	01	-11	-13	-13	-29**
Two dimensional space		-02	-05	06	02	-10	-34**
Three dimensional space	(-55)**	11	09	-05	01	03	-28*

[a] Decimal points omitted.

*$p < .05$,

**$p < .01$, directional hypothesis

Table 8

Means and Standard Deviations for Test Scores and Choice RT, Studies 1, 2, and 3

	Study 1 High school 1973		Study 2 College 1975		Study 3 High school 1975	
Measures	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Verbal comprehension	13.1	6.0	19.4	4.2	18.6	4.0
Numerical ability	15.8	7.8	27.5	8.3	26.2	8.0
Numerical reasoning	8.2	3.8	13.8	2.4	12.5	2.9
Verbal reasoning	13.0	5.9	18.6	3.8	17.4	4.8
Symbolic reasoning			16.0	6.3	14.0	6.9
Progressive matrices			26.5	4.8	24.0	6.4
Maze tracing			21.4	7.7	26.4	7.0
Hidden figures	3.4	3.1	11.2	6.3	10.2	6.9
Two dimensional space			16.3	5.0	18.6	4.1
Three dimensional space	(19.2)	(9.8)	15.3	4.4	15.7	5.0
One finger CRT initial	338.0	83.7	331.2	43.8		
One finger CRT terminal			321.3	37.1		
Two finger CRT initial			285.4	27.0	303.8	71.2
Two finger CRT terminal			270.3	28.6	288.2	42.8

Table 9

Predicted Decrements (Regression Coefficients) in Test Score Associated with Ten Millisecond Increase in Choice Reaction Time[a]

	Choice RT one finger			Choice RT two finger			
	High school 1973	College 1975		College 1975		High school 1975	
		Block 1	Block 6	Block 1	Block 6	Block 1	Block 3
Verbal comprehension	.33	--	--	.17	--	--	--
Numerical ability	.45	.11	.04	.58	--	.04	.35
Numerical reasoning	.23	.05	.05	.20	--	.06	.08
Verbal reasoning	.23	.03	.25	--	--	--	--
Symbolic reasoning	x	.11	.29	.19	--	.08	.50
Progressive matrices	x	.05	--	.20	.03	--	.40
Maze tracing	x	.21	.39	.09	--	.28	.46
Hidden figures	.14	.04	--	.25	.29	.13	.47
Two dimensional space	x	.02	.07	--	--	.06	.33
Three dimensional space	(.64)	--	--	.08	--	--	.33

[a] x indicates no data available and -- indicates relationship reversed from expectation.

EAC
REPORTS

EDUCATIONAL ASSESSMENT CENTER

University of Washington, 1400 Campus Parkway, PB–30, Seattle, Washington 98195 (206) 543–1170

Instructional Evaluation

Educational, Psychological, and Survey Research

Test Administration, Construction, Scoring, and Analysis

Educational Assessment Center

University of Washington

March 1977

77-10

Information-Processing Correlates of the Raven Progressive Matrices Tests

Clifford E. Lunneborg

Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977. Research reported was supported jointly by grants to The University from the National Institute of Education (C. E. Lunneborg, Principal Investigator) and The National Institute of Mental Health (E. B. Hunt, Principal Investigator).

Educational Assessment Center Project: 262

Information-Processing Correlates of the Raven Progressive Matrices Test

Clifford E. Lunneborg

University of Washington

Among the techniques for assessing intellectual ability, at least in the English speaking world, a major emphasis has been the measuring of what has come to be termed verbal ability or verbal intelligence. Classically indexed by paper and pencil tests of vocabulary--tests of the recognition of synonyms or antonyms--this ability has been widely used as an indicator of academic promise. Indeed, such measures have been so taken as nearly synonymous with academic intelligence, at least, that the study of verbal ability seemed a natural place for us to start to look for relations between psychometric measures of cognitive abilities on the one hand and measures of information processing as they have been developed, theoretically and empirically, by modern cognitive psychology on the other. We found quite interesting relations and have since reported on the information processing characteristics of the "high verbal" (Hunt, Frost and Lunneborg, 1973, and Hunt, Lunneborg and Lewis 1975).

Verbal tests, of course, are not the only widely used assessments of intellectual ability. In particular, recent years have seen a growth of interest in measures which, at least on the face, would appear to be less culture-bound. One's vocabulary, after all ought to reflect on one's particular learning history

---

Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977. Research reported was supported jointly by grants to The University from the National Institute of Education (C. E. Lunneborg, Principal Investigator) and The National Institute of Mental Health (E. B. Hunt, Principal Investigator).

as well as on one's "ability". Among the measures which have been proposed as less history-bound, less "crystallized" to make Cattell's distinction between crystallized (learned) and fluid (ability) intelligence, is the Raven Progressive Matrices test developed in Great Britain. The test involves non-verbal problems, figural reasoning problems of a kind not deliberately taught in schools. What information processing characteristics might be associated with those who do well at the Raven task? How might they differ from those characteristic of the "high verbal?" In one of our recent "marathon studies"--studies in which the same group of subjects participate in a large number of psychometric and laboratory assessments,--we had an occasion to begin to look at this question.

Method

Subjects

Sixty-three University of Washington freshmen (36 women and 27 men) were recruited fall 1974 from those who had completed the battery of tests administered high school juniors by the Washington Pre-College (WPC) Testing Program. These volunteers were selected so that the joint distribution of the WPC Verbal and Quantitative Composite scores in the sample reflected the distribution for the entire entering class. Participants were paid and served as subjects over an academic quarter.

Psychometric Measures

A vocabulary score (VOC) was calculated as the average of the high school administered WPC Vocabulary test, a test of recognition of synonyms, and of the Synonym-Antonym section of the Terman Test of Concept Mastery. The Terman Test was administered during the project and both sets of scores were standardized within the sample before averaging.

The Raven score resulted from the administration of Set II of the Advanced Progressive Matrices. Participants were given 50 minutes to respond to the 36 items, pacing their own performances. Score is the number of items answered correctly.

As an additional source of information about cognitive functioning, as viewed psychometrically, the Wechsler Adult Intelligence Scale (WAIS) was administered individually to members of the sample. Verbal, Performance and Full Scale IQs were established. The Raven was administered at the beginning of the project period and the WAIS at the conclusion of all other participation.

All of these measures of intelligence are, in some sense, measures of power. Because our earlier researches had focussed attention on speed of processing we decided to include, for the present analyses, some speeded indicators, among those intellectual measures. In the course of two laboratory tasks, response time measures were obtained and these will be treated here together with the psychometric measures of intelligence. Individually, participants were presented, one item at a time, Set I of the Advanced Progressive Matrices. This was done some weeks after group administration of Set II and involved 12 problems not in the larger set. Time to respond to each item--as well as choice of response--was recorded for each item. A principal components analysis of these data suggested that the time to respond to earlier and to later items in the series were separable. Two scores were used in this study: RAVEN TIME 1, the average of the standardized times to solve problems 1 and 2 and RAVEN TIME 2, the average of the standardized times to solve problems 8 and 9.

A second laboratory task called for learning a list of statements which associated each of a set of personal names either with occupations (Ruth is a lawyer) or locations (John is in the pool), or which associated locations with activities (If you are in the park you can hear the chimes ring). A week later, following a recall test of mastery of these "facts", a series of True/False questions was presented. The questions were sush as to require the retrieval of 1, 2, 3 or 4 of the "facts". Time to answer each question was recorded and two scores of interest derived. Verbal Problem Solving Slope (VERBAL SLOPE) and Verbal Problem Solving Intercept (VERBAL INTERCEPT) both resulted from linearly regressing the time to correctly respond to an item on the number of "facts" required by the item. These nine measures, including the Raven, are the "criteria" for the present study.

Information Processing Measures

As a first laboratory task, participants fixated a midpoint on a CRT display and responded by key press to the onset of a character to the left or right of that midpoint. In counter-balanced order students participated under two conditions. In the one-finger condition, the index finger of the preferred hand was used to depress one of two keys. In the two-finger condition keys under left or right index fingers were used to signal detection of stimuli to left or right respectively. Using median response times, separately for presentations to the left and right, for the first and last (sixth) blocks of 100 presentations under each condition, two principal components to these choice reaction times were identified. As a result, two composite measures were used in this study: CRT1 was the average of the standardized times to correctly respond for left and right presentations, first and last block under the one-finger condition. CRT2 was similarly defined for the two-finger condition.

In a visual temporal order judgement task students signalled by left or right hand key-press whether the first of two visual displays to a CRT screen appeared to the left or right of the screen. Onset asynchrony was counterbalanced between the sides and was either 0, 15, 30, 45, or 60 milliseconds. Data treated here are based on the proportion of correct responses under the different conditions. LFIRST was the proportion of correct responses when the left stimulus led the right by either 30 or 45 milliseconds. RFIRST was identically defined for right stimulus led trials. VISUAL SLOPE and VISUAL INTERCEPT described the linear regression of proportion correct judgements on the stimulus onset asynchrony interval (averaging over right and left led trials).

In a dichotic listening task pairs of speech sounds (ba, da or ga) or pairs of non-speech sounds (buzz, hiss or tone) were presented separately to the left and right ears with one sound leading the other by 50 or 100 milliseconds. Participants were asked to identify which input occurred first and the number of correct responses recorded for each ear by lag condition. LNS and RNS were the number of correct responses to non-speech stimuli with the sound to the left or right ear leading, respectively. LSPCH and RSPCH were correspondingly derived from the speech sound trials.

Students also participated in a delayed auditory feedback task, reading aloud a page of what was English prose, the words having been randomly rearranged. Each participant read under three conditions; the reading was played-back through earphones with zero, 150 or 300 milliseconds delay. READ was the time taken to read the page with no delay in feedback. READEL was the average of the standardized times taken to read under the two delay conditions. ERSP and ERIN were counts of errors made in the reading under the delay conditions, the first a count of speech errors (stuttering, substituting a non-word for a word, etc.) and the latter a count of inversions, occasions on which word order was changed in the reading. (Counts of two other kinds of errors were also made, though not included in this study. Number of words changed to other words loaded the same component as the number of speech errors in the analysis of measures collected in this task while the count of words omitted was not systematically related under the two delay conditions.)

Finally, during the delayed auditory feedback experimental session, participants were asked to read back aloud each of ten visually presented words substituting L's for R's and R's for L's in the reading. RTOL was the median conversion latency across the ten words.

Results

Table 1 reports the correlations of the information processing measures with the psychometric scores. The choice response times were not significantly correlated with any of the psychometric measures. Accuracy at the visual temporal order judgement task, notably when the leading stimulus is to the left visual field, correlated with WAIS performance IQ's and speed in responding to the RAVEN items. A similar pattern of correlations was obtained for LNS, the accuracy measure for those dichotic listening trials involving non-speech stimuli and for which the stimulus to the left ear led. RNS, accuracy in identifying those instances where a non-speech signal to the right ear preceded one to the left, correlated--in contrast--with the two verbal measures of intelligence (VOC and WAIS-V). Accuracy in distinguishing the order of speech sounds, right ear leading left ear (RSPCH), was the only measure with a significant correlation with the Raven score. When the speech signal was fed first to the left ear (LSPECH), accurate identification was related to quickness in the verbal problem solving task.

Table 1

Correlations Between Psychometric Measures and Performance on Information Processing Tasks

	VOC	RAVEN	WAIS-V	WAIS-P	WAIS-F	RTIME1	RTIME2	VERB S	VERB I
CRT1	10	-03	04	-22	-11	-07	-10	10	05
CRT2	-10	-07	10	-01	05	01	-21	-14	-25
LFIRST	25*	08	16	37**	30*	-35**	-47**	-10	02
RFIRST	03	11	-27	-14	-24	17	-05	-06	-12
VIS SLOPE	-02	-04	19	-29	-04	07	13	-10	09
VIS INTER	-08	-00	10	-01	04	00	00	04	-01
LNS	12	06	24	57**	47**	-15	-33**	-17	-04
RNS	43**	01	31*	22	33**	08	01	21	-10
READ	-44**	-07	-30*	-04	-23	-26	-26	09	28*
LSPCH	11	15	-05	13	04	09	02	-26*	-26*
RSPCH	15	33**	-16	06	-07	-01	-13	-13	-15
READEL	-31*	-02	-09	-13	-14	-27	-21	-13	36**
ERSP	-29*	-02	-24	-13	-23	-12	-09	-07	25*
HRIN	18	06	35	05	22	-25	-10	46**	12
RTOL	-54**	-21	-43**	-20	-40**	04	-02	02	12

Decimal points omitted. $^{*}p \leq .05$ $^{**}p \leq .01$

Speed in reading through a stream of randomly ordered words, with or without the distraction of a delayed auditory feedback (READ and READEL), was related to better vocabulary performance as well as greater speed in responding to the verbal problems. Speech errors (ERSP) in the delayed auditory feedback task had these same high correlations, while inversion errors (ERIN) slowed the solution of verbal problems requiring the retrieval of larger numbers of facts. Finally, speed in accomplishing the R to L and L to R speech conversions was the strongest correlate of the two verbal intelligence measures.

There was fair consistency in the correlates of the verbal measures, then, and the two, as shown in Table 2, were well related to each other. The Raven, however, was not highly correlated with any of the other psychometric measures. The problem solving speeds, both for the Raven and Verbal problems, were related to WAIS scores, particulary to those contributing to the Performance IQ.

To attempt to gain further insight, those information processing measures which correlated significantly with the psychometric scores were employed as potential multiple predictors of those psychometric scores. The results of the multiple predictor selections are given in Table 3. While 36% of the vocabulary variability in this sample was associated with these predictors only 14% of the Raven score differences could be accounted for. This greater predictability of verbal comprehension replicates a finding of one of our earlier studies, (Lunneborg, 1977.)

Speed in the R to L conversion task and accuracy in detecting that a non-speech stimulis to the right ear preceded one to the left were the major determiners of verbal test performance. Accuracy when the non-speech stimulus occurred first on the left side and accuracy in identifying that a visual stimulus occurred first to the left side account for an appreciable amount of the variability in

Table 2

Correlations Among Psychometric Measures

	RAVEN	WAIS-V	WAIS-P	WAIS-F	RTIME1	RTIME2	VERB S	VERB I
VOC	26	62	01	41	-12	-04	28	-16
RAVEN		-26	-03	-19	-17	-02	-11	21
WAIS-V			41	88	-26	-22	-21	-55
WAIS-P				79	-48	-25	-43	-22
WAIS-F					-43	-29	-38	-51
RTIME1						54	17	-08
RTIME2							17	01
VERB S								-03
VERB I								

WAIS performance IQ's and speed in solving the more difficult Raven problems. Raven performance itself, however, could only be linked to accuracy in noting that of two speech stimuli the one to the right ear was earlier in time. Relative resistance to distraction--reflected in relative resistance to the effect of delayed auditory feedback and superior accuracy in preserving the order of input of two speech signals when the first was heard in the left ear--contribute to speed in solving verbal problems involving retrieval of one to several learned facts. The two contributors to the prediction of WAIS full scale IQ's, LNS and RTOL, are those associated with the two WAIS components.

Discussion

This study, unfortunately, has taught us little about determiners of Raven performance. Just as the Raven was uncorrelated in this sample with other measures of intelligence so was it much less well predicted from the several information processing measures. On the other hand, we again found that the verbal performance investigated earlier and exemplified here by the vocabulary score, was highly predictable from measures of information processing. While preserving order information, noted in our earlier reports, continued to be significant, it was the ability to overcome linguistic convention, in the R to L conversion task, that was here the major determiner of those verbal scores. Correlates of the WAIS-V scores, partially dependent upon a vocabulary test, were essentially the same as those for the vocabulary measure itself.

IQ's derived from the non-verbal or performance subtests of the WAIS, quite interestingly, were well predicted by a pair of information processing measures, LNS and LFIRST, that tap ability to maintain the order of input of non-verbal stimuli, either auditory or visual. Although Raven scores were not correlated

Table 3

Multiple Predictors of Psychometric Test Performance

Criterion	Predictor	R^2	Increment
VOC	RTOL	289	
	RNS	347	058
	LFIRST	362	015
RAVEN	RSPCH	110	
	RTOL	141	031
WAIS-V	RTOL	181	
	RNS	204	023
WAIS-P	LNS	329	
	LFIRST	356	027
WAIS-F	LNS	221	
	RTOL	336	115
RTIME1	LFIRST	124	
RTIME2	LFIRST	222	
	LNS	248	027
VERB INTER	READEL	128	
	LSPCH	171	042
VERB SLOPE	ERIN	207	
	LSPCH	217	010

Decimal points omitted

with WAIS-P IQ's, these same two indices of accuracy of judgement of the temporal order of non-verbal stimuli also contributed to the prediction of the time required to solve Raven problems. Finally, there is the suggestion that verbal problem solving proficiency (as distinct from the verbal comprehension ability measured by a vocabulary test) may be grounded in an ability to accurately maintain the order of speech elements. These two findings--that, at least some, measures of non-verbal intelligence or problem solving are linked to accuracy in recovering the order of non-verbal stimuli and that verbal problem solving efficiency is linked to accuracy in recovering the order of verbal stimuli--represent, we believe, this study's contribution to the further development of relationships between psychometrically evaluated abilities and measures of information processing efficiency. Both relationships offer testimony to the potential productivity of research in this area.

What may underlie the relatively poor showing of the Progressive Matrices test? Recent studies have suggested first that the Raven items may not be as homogeneous in their cognitive demands as was earlier supposed (Hunt, 1974) and second that there may be important differences between groups of respondents, as between men and women (Chiang and Atkinson, 1976), in the relations among information processing measures. Further understanding of the Matrices approach to intellectual assessment may depend, therefore, upon the successful clustering of items and/or of test-takers. We are working on both tasks at Washington now.

References

Chiang, A. and Atkinson, R. C. Individual differences and interrelationships among a select set of cognitive skills. _Memory and Cognition_, _4_, 661-672, 1976.

Hunt, E. B. Quote the Raven? Nevermore! In L. Gregg (Ed.) _Knowledge-Cognition_, Hillsdale, N.J.: L. Erlbaum Assoc., 1974.

Hunt, E., Frost, N. and Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.) _Advances in learning and motivation_, Vol. 7. New York: Academic Press, 1973.

Hunt, E., Lunneborg, C. and Lewis, J. What does it mean to be high verbal? _Cognitive Psychology_, _7_, 194-227, 1975.

Lunneborg, C. E. Choice reaction time: What role in ability measurement? _Applied Psychological Measurement_, in press, 1977.